



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Common Ground

Student and Teacher Assessment Criteria

Graeme Addinsall

Examensarbete:	15 hp
Program och/eller kurs:	LAU925:2
Nivå:	Grundnivå
Termin/år:	Vt/2012
Handledare:	Davoud Masoumi
Examinator:	Christian Bennet
Rapport nr:	VT12-IPS-04 U/V ULV LAU925

Abstract

Examensarbete:	15 hp
Program och/eller kurs:	LAU 925:2
Nivå:	Grundnivå
Termin/år:	Vt/2012
Handledare:	Davoud Masoumi
Examinator:	Christian Bennet
Rapport nr:	VT12-IPS-04 U/V ULV LAU925
Nyckelord:	Assessment, national tests, formative assessment, transparency.

The aim of this study is to identify the central assessment criteria that students themselves use to assess and grade texts written in English and compare them to the assessment criteria provided by the National Agency (Skolverket) in their teacher's instruction booklet related to the National Test of English. To do this, the fourteen example-texts provided by Skolverket in their teacher's instruction booklet are assessed by students in focus groups. Students themselves group their responses to the texts in terms of strengths and weaknesses. They also give each text a grade. Further classification is then done according to the two categories – Content and Language – and the eleven related text analysis criteria ('bedömningsfaktorer') found in the teacher's instruction booklet. Skolverket's comments relating to each text, which are in the form of free-flowing text, are also broken down and classified according to the same criteria. A comparative analysis is then done relating to the categories, criteria and grades.

This study finds that the students and Skolverket employ a common set of assessment criteria in their responses. Moreover, there is a strong correlation between the relative weights given to each criterion over the span of the fourteen texts. However, it is also shown that students have more focus on language over content in their assessments than Skolverket do and that while responses from Skolverket focus on the strengths in the texts, student responses focus on weaknesses. Further, using Skolverket's whole grades as a benchmark, the students grade sixty percent of the texts the same as Skolverket. In addition, the data shows that the accuracy of student grading diminishes sharply at higher levels. There is an inverse relationship; the higher the grade the lower the accuracy.

Foreword

This study offers me the chance to investigate an area which has been of interest to me for some time. I have worked as an English teacher in a junior secondary school in Sweden for eight years. During this period I have been responsible for the coordination and correction of the national tests for numerous grade 9 classes. As a result, I have regularly read and analysed the relevant material related to the tests, both for students and teachers.

Over the years, I have also noticed much more focus in the wider pedagogical debate on formative assessment practices. The prominence of this theme can be seen by the sheer number of professional development seminars and programmes which are now available. Its influence is also reflected in the official literature provided by the Swedish National Agency for Education (Skolverket). In my view, its theoretical footprint is quite explicit in Lgr11. That is to say, formative assessment is built into the new course plans and syllabuses. This raises the question of how far we have already come. To what extent do teachers and students speak the same language in regard to the assessment criteria they use?

This study is an attempt to empirically document this question using the national test of English as a platform and the National Agency's recommended text analysis criteria as a benchmark.

I would like to thank all the students who took part in this study. I would also like to express my gratitude to their teachers who made time for me despite their busy schedules, as well as my supervisor, Davoud Masoumi, for his support and advice.

Table of Contents

1	Introduction	6
1.1	Aim	6
1.2	Research Questions.....	7
2	Literature Review and Previous Research.....	7
2.1	The National Test	8
2.2	The Grading Scale	9
2.3	Interpreting the Assessment criteria	9
2.3.1	Teachers	10
2.3.2	Students	12
3	Method.....	15
3.1	Data Collection and procedure	15
3.1.1	Focus Groups.....	15
3.1.2	Participants	16
3.1.3	Assessment criteria.....	16
3.1.4	Developing the Exercise.....	18
3.1.5	Procedure.....	18
3.2	Data Analysis.....	19
3.3	Reliability, Validity and Generalizability.....	19
3.4	Ethical Issues	20
4	Findings.....	20
4.1	Student Responses	21
4.1.1	Categories: Content and Language	21
4.1.2	Assessment criteria used by Students.....	21
4.1.3	Focus on Strengths and Weaknesses: A Comparison	22
4.2	Categories: A comparison	22
4.1.3	Categories.....	23
4.1.4	Assessment criteria.....	23
4.1.5	Focus on Strengths and Weaknesses	24
4.1.6	Focus on Strengths and Weaknesses within each Criterion	25
4.2	Grades: A Comparison	26
4.2.1	Overview	26
4.2.2	Various Frames of Reference for Comparison.....	27
5	Analysis	29
5.1	Common Assessment Criteria	29
5.2	Divergence.....	30
5.3	Grading	31

5.4	An Attempt at Synthesis	31
6	Conclusions	34
7	References	36
8	Appendix	40
8.1	Attachment 1.....	40
8.2	Attachment 2.....	42
8.3	Attachment 3.....	43
8.4	Attachment 4.....	46

1 Introduction

Both Lpo94 and Lgr11 are based on democratic principles which promote individual rights and tolerance (Skolverket, 2006; 2011a). The values embraced aim to foster both inclusivity and openness. The focus on democratic ideals has also had broad political and economic implications over recent years, not least being the decentralization of the government school system from state to municipal authorities, the promotion of government-funded independent schools and the introduction of the right of students to freely choose the school they would like to attend (Sweden. se, 2009).

Democratic impulses have also prompted changes within the classroom. Technology is playing an increasing role in schools and aiding networking and even international contacts (Stiernstedt, J., 2012). The syllabuses in both Lpo94 and Lgr11 are outward looking, with a face toward the wider community and the world at large. However, democratic changes have also taken place at a more individual level, with the aim of clarifying the goals and assessment criteria related to the learning process. Reforms in the law and statutory regulations have paved the way for such changes as the implementation of IUP (Individuell utvecklingsplan) which means that it is now compulsory for all students to have an IUP, which is evaluated and updated each term. Together with the written assessment 'skriftlig omdöme' the IUP is meant to provide both a clear description of the goals fulfilled up to that point in time as well as specify goals for the future and outline a plan for how to achieve them (Skolverket, 2009a).

Another recent addition aimed at promoting clarity for students and teachers is pedagogical planning 'pedagogisk planering', simply called 'planning' by Skolverket in recent publications (Skolverket, 2011c; Skolverket, 2011b). Skolverket advises teachers to plan their teaching according to a number of criteria including clearly defined and appropriate content, specified forms of presentation, and active student participation (Skolverket a, 2011).

These pedagogical developments have multiple purposes; however one common theme that can be gleaned from them is a general move toward transparency, not only for teachers, but also for students and guardians. The discussion that takes place at the parent-teacher meeting (utvecklingssamtal) between the student, guardians and teacher about what the goals are and future plans to reach them; or between teacher and students in the formulation of pedagogical planning, is of vital importance for clarifying fundamental aspects of the learning process for all concerned. The aim of this study is to provide a glimpse of how far we have come in terms of achieving transparency with specific reference to how students assess and grade texts written in English.

1.1 Aim

The aim of this study is to identify the central assessment criteria that students themselves use to assess and grade texts written in English and compare them to assessment criteria provided by the National Agency (Skolverket) in their teachers' instruction booklet related to the national tests of English.

1.2 Research Questions

This study focuses on the following questions:

- What are the central criteria that the students employ in their assessment?
- To what extent do student assessment criteria reflect the criteria provided by the National Agency (Skolverket) in their teacher's instruction booklet related to the National Test of English
- Irrespective of the criteria, how do the grades set by the students compare to those provided by Skolverket?

2 Literature Review and Previous Research

Assessment has long been a central area of investigation of pedagogical research. During more recent times, the debate over formative and summative assessment has played a significant role in the debate. Helen Korp (2003) includes a historical overview of the debate in *Kunskapsbedömning – hur, vad och varför* and raises the issue that is central to this investigation, namely the relationship between student and teacher assessment. Lpo94 is built on a view of knowledge that has four pillars; facts, understanding, skills and passive knowledge. It constitutes a further move away from a predefined curriculum to a decentralized model where interpretation of national goals is required at the local level (Skolverket, 2009). How these goals are interpreted in the staffroom and the classroom is therefore of utmost importance for student learning.

It also raises the issue of what the assessment criteria are and how they should be applied. Frederiksen and Collins (1990) stress the importance of students having a sound understanding of the assessment criteria. Cunningham (1998) views it as crucial that all criteria are fully defined and made available to students prior to a course starting. Eisner (1991) raises a number of problems with this position claiming that doing so limits the creative nature of the teaching process. Others take the debate one step further. Shepard (2000), for instance claims:

The mere provision of explicit criteria will not enable learning in all the ways desired if they are imposed autocratically and mechanically applied. For the intended benefits to occur, self-assessment has to be a part of more pervasive cultural shifts in the classroom. Students have to have the opportunity to learn what criteria mean (surely not memorize them as a list), be able to apply them to their own work, and even be able to challenge the rules when they chafe (p. 61).

Since the 1960's the debate over assessment and more specifically, its role in the learning process has gained momentum (Korp, 2003, p. 79). As noted above, summative and formative assessment has become one of the debate's focal points. It is difficult to find a more eloquent description of the two forms than the quote by Robert Stake (1999), cited by (Korp, 2003).

Formative assessment is when the chef tastes the soup and summative assessment is when the guests do it (p. 77). [My translation]

Formative assessment is essentially a teaching method which is employed to promote learning. Summative assessment is a tool used to determine student outcomes at a given point in the learning process. They have very different functions. Summative assessment is not inherently concerned with the transparency of assessment criteria. Its role is to assess student performance in relationship to set criteria. The extent to which students are aware of the assessment criteria is irrelevant to the actual test as such. After the test is completed, the

criteria are applied and a result is arrived at which is seen as a measure of competency at that point in time. The national tests are summative in this sense. Student performance is assessed in relationship to national goals in each respective subject. The extent to which specific assessment criteria are transparent for students is irrelevant to the assessment of student performance. Formative assessment has its focus on the process of learning rather than a measurement of performance. It emphasizes involving the student in the assessment process via such methods as self-assessment, peer-assessment, feedback loops and a variety of models of documentation including portfolio which are designed to engage students in their own learning and help them reflect on paths for development. As such, the assessment process is primarily viewed as a pedagogical tool rather than a means of measuring competency. Nevertheless, research into formative assessment methods indicates that they can lead to improved student performance (Shepard, 2000).

One could say that the formative model aims at transparency at an individual level. It aims at reflection and analysis. It allows for and encourages multiple sources of input in terms of assessment, which is in stark contrast to the traditional teacher-dominated model. It enables assessment criteria to be created by teachers and students as part-and-parcel of the learning process. Not least, formative assessment lends itself to the development of common assessment criteria for students and teachers. The importance of shared criteria is highlighted by Shepard (2000) when he says, “The features of excellent performance should be so transparent that students can learn to evaluate their own work in the same way that their teachers do” (p. 60).

My objective is to investigate the extent to which such common ground exists by eliciting student assessment criteria relating to specific texts. As such, this study is a measure of formative assessment in a summative form. How similar are the criteria used by students and teachers when the criteria are not formally imposed or defined? In other words, what are the students own “internal criteria?” (Korp, 2003, p. 84)

Lpo94 has been with us 17 years. This is the last year that it will be used as the basis for the obligatory Swedish education system. Therefore, this is also the last year that students will be assessed in accordance with Lpo94’s assessment criteria. This offers a unique opportunity to research issues relating to student perceptions of assessment and compare them to those advanced by the Swedish National Agency for Education and Lpo94.

2.1 The National Test

One of the central aims of the national tests is to promote “...fair, standardized and reliable assessments” (Skolverket 2005, p. 26). At present in Grade 9 there are national tests in Swedish, Swedish as a Second Language, Mathematics, English, Biology, Chemistry and Physics. In order to support teachers in their assessment of the tests, Skolverket provides teachers with example material as well as answer sheets for the relevant exercises. The English test is divided into three sections: Part A: Oral Interaction and Production, Part B: Receptive Skills, and Part C: Written Production. Both the oral interaction and written production sections include examples of student performance. An assessment of each example is also provided.

This study relates to Part C of the national test of English for Grade 9. The focus of the study requires that I use the text examples from Part C and the related comments provided by Skolverket in the Teacher’s booklet, *Bedömning och exempel* (Skolverket, 2008). This is

because the study involves both collecting and analysing student responses to texts as well as mapping out the responses provided by Skolverket and then comparing them to those given by the students. By using material in the Teacher's Booklet, all of these requirements can be met.

Using these texts allows me to employ Skolverket's recommendations as my benchmark. These are the criteria that teachers use to help them hone their assessment skills. The text analysis, comments and grading provided are therefore as objective as is possible to attain in the sense that they encompass the official interpretation, advanced and supported by the National Agency. The debate that usually surrounds assessment is commonly focused on teacher assessment. If I had used other texts for my study, I would have needed to resolve the problem of what constitutes a legitimate or acceptable analysis and assessment of the texts. By using Skolverket's material, I avoid the need to deal with this issue and this means I can focus on the task at hand, which is analysing student responses and comparing them to those provided by the national Education Agency.

This doesn't mean however that there are no problems relating to how the texts are classified or interpreted. I will explore this issue further in the Method section.

2.2 The Grading Scale

Lpo94 uses the following system for allocating grades.

The Awarding of Grades in Basic Compulsory School	
Grade	Interpretation
-	Has not yet attained all goals in the subject
Pass (G)	Has attained all goals in the subject
Pass with distinction(VG)	Has attained all goals in the subject and satisfies the criteria for the award of "pass with distinction"
Pass with special distinction (MVG)	Has attained all goals in the subject and satisfies the criteria for the award of "pass with special distinction"

(Skolverket 2005, p. 14)

2.3 Interpreting the Assessment criteria

Lpo94 includes a syllabus for each subject which lays out the goals to strive for and the goals to be fulfilled at the end of grade 5 and grade 9. The goals to be fulfilled represent the minimum expectation of performance. In grade 9 they all need to be met in order to achieve a pass in the subject (in the absence of special considerations). The criteria for VG (pass with distinction) and MVG (pass with special distinction) are also included for Grade nine. The national goals that specifically relate to writing are the following:

Mål att sträva mot		
<ul style="list-style-type: none"> • utvecklar sin förmåga att använda engelska för att kommunicera i (tal och) skrift • utvecklar sin förmåga att uttrycka sig varierat och säkert i skrift för att berätta, beskriva och förklara samt motivera sina åsikter • utvecklar sin förmåga att delta aktivt i (samtal och) skriftlig kommunikation, uttrycka sina egna tankar på engelska samt uppfatta andras åsikter och erfarenheter 		
Mål att uppfylla		
Godkänt	Väl Godkänt	Mycket väl godkänt
Eleven skall kunna begära och ge information i skrift samt berätta och beskriva något	Eleven skall kunna skriva varierat och med sammanhang samt kommunicerar skriftligt vid informationsutbyte och sociala kontakter, ställer och besvarar frågor och anpassar sitt språk till några olika mottagare	Eleven skall kunna uttrycka sig varierat i skrift och anpassar framställningen till några olika syften och mottagare

(Skolverket, 2000, para. 3)

The question that teachers are faced with is how to apply these criteria in practice. How can student performance be reliably assessed using such open-ended criteria as those defined in the national standards? The problem is multifaceted; however any solution needs to accommodate two major themes. Firstly, it is necessary to ensure that teachers have a common understanding of what the national goals mean and that they apply the assessment criteria consistently and equitably in their assessments. Secondly, it is necessary to ensure that the students also understand basis on which they are assessed. The first point directly relates to the ongoing debate about fair and reliable assessment. The second raises issues which are perhaps more related to pedagogy and didacticism.

2.3.1 Teachers

Skolverket explained what is meant by fair and equitable assessment in 2004.

En rättvis betygsättning innebar att det betyg en elev fått i ett visst ämne eller i en viss kurs ska visa elevens kunskaper och färdigheter i det som ska ingå i kursen enligt kursplanen och som motsvarar betygsriterierna för betyget. Med likvärdig bedömning menas att måttstocken för bedömningen är detsamma för alla elever. Ett betyg i en klass ska motsvara ett likadant betyg i en annan klass (Skolverket, 2007, p. 9).

Meeting this demand presupposes that certain conditions are met.

Grundläggande för en rättvis och likvärdig betygsättning är att de som sätter betygen och som utfärdar betygsdokumenten, dvs, lärare och rektorer, har tillräckliga kunskaper (Skolverket, 2007, p. 9).

However, just knowing what the criteria are, is according to Skolverket, insufficient to ensure fair assessment practices.

Men enbart att känna till reglarna räcker inte, det krävs också ett aktivt lokalt kursplanearbete där betygsriterierna tolkas och anpassas utifrån det stoff och de undervisningsmetoder som används. Det förutsätter en bred diskussion mellan lärare såväl inom som mellan olika skolor om hur man ser på måloppfyllelse och kunskapskriterier. Samrådet krävs för att trygga en likvärdig och rättvis betygsättning (Skolverket, 2007, p. 9).

The national tests are not meant to define final grades; rather their aim is to assist teachers in their final assessments. “In this process, national tests are only one of several assessment components.” (Skolverket, 2005, p. 25) As such, divergence between national test results and final grades is to be expected. There are a number of reasons why final grades may differ from national test results (Skolverket, 2007). However, despite this fact, the tests themselves provide teachers and students with standard tests which are assessed using a common set of criteria for each subject. Given these conditions, it is perhaps not surprising that, as previously noted, one of the aims of the tests is to ensure fair, standardized and reliable assessment practices (Skolverket, 2005).

In 2000 Skolverket initiated a national study which focused on assessment. Its findings pointed to large discrepancies in how the criteria are interpreted and applied (Skolverket, 2007). This contributed to a further study being undertaken, *Rapport 300* which examined the relationship between final grades in grade 9 and national test results in Swedish, English and Mathematics. Its aim was to contribute to the debate by comparing national test results to final grades. The study concluded that more remains to be done to secure fair assessment practices.

Det finns skolor där en mycket stor andel elever får ett *högre* slutbetyg än provbetyg och det finns skolor där stor andel får ett *lägre* slutbetyg än provbetyg (Skolverket, 2007, p. 6)[Original italics].

As a whole, final grades are considerably higher than test grades. The reason for the discrepancy is unclear, however a few possible explanations are proposed. Firstly, that it is a result of extra support provided to the students after doing the test. Secondly, that teachers find it difficult to fail students, or they base their grading on criteria that aren't linked to the national criteria. Thirdly, that teachers in different schools interpret the criteria in different ways (Skolverket, 2007).

Clearly the first point does not threaten fair assessment practices. On the contrary, it can be seen as a positive result of the test and a good example of why final grades can differ from test scores. On the other hand, if point two plays a role then fair assessment practices are threatened. Point three is interesting as interpretation of the assessment criteria is not only an accepted component of the assessment process, but a requirement; nevertheless, it can compromise fair assessment practices. Skolverket recognises this point.

Det här är den kanske viktigaste och mest troliga förklaringen som kan påverka likvärdigheten i betygsättningen även om den ligger inom ramen för betygsbestämmelserna (Skolverket, 2007, p. 43).

Here we arrive at the central issue – interpreting the criteria. Skolverket points out that interpretation of the criteria isn't done by just anyone. The system is based on the fact that teaching professionals have a “...common understanding of the concepts used in the grading criteria and that they can also assess students on similar grounds” (Skolverket, 2007, p. 41) [My translation]. According to Skolverket,

Lärarna är suveräna i sin betygsättning och det finns ingen som kan överpröva lärarnas tolkning av nationella mål och kriterier så länge de håller sig till dessa och inget annat vid betygsättning (Skolverket, 2007, p. 43).

In an earlier report from Skolverket which is aimed at the international audience, similar discrepancies in assessment are also noted. Commenting on the results comparing final grades with the national test results for Swedish in 2004, the report states that in some schools 40 percent of students received a higher final grade than for the national test.

Such a skewed difference is a dilemma in a school system when, at the local level, there is a large degree of freedom for interpretation and assessment of syllabus and grading criteria (Skolverket, 2005, p. 28).

The results of a current audit undertaken by Skolinspektionen reinforce the concerns noted above. Selected national tests are to be reassessed using a control group of assessors. Approximately 35,000 national tests from 750 schools are to be scrutinised annually, over a period of three years. The tests range from Grade 3 to the final year of secondary college. Two of the three reassessments have already been completed. The subjects involved in the audit are Swedish, Maths and English (Skolinspektionen, 2011). In summarising the results for the year's audit, Skolinspektionen comments,

Resultatet av omrättningen visar att det generellt sett finns omfattande och stora avvikelser mellan ursprungs rättarens bedömning och Skolinspektionens bedömning för vissa delprov. Detta gäller framför allt de delprov där eleven ska avge sitt svar i form av en uppsats (Skolinspektionen, 2011, p. 4).

According to an article in Göteborgs-posten 14 November, 2011,

I nästan hälften av de insamlade proven från 62 gymnasier åren 2009 och 2010 satte de omrättande lärarna ett annat betyg, i huvudsak ett eller flera betygssteg lägre, än ursprungs rättaren. (Isemo, 2011).

So here we see that almost half of the teachers in the schools have set a grade for the same test *at least* one grade higher than Skolverket's control assessors [My italics]. However, an even more surprising statistic is also revealed.

Mest anmärkningsvärt är att Skolinspektionens lärare för var nionde elev satte betyget IG i fall där berörda skolors egna lärare hade bedömt provet vara värt ett MVG (Isemo, 2011, para. 4).

Skolinspektionens project leader, Arletta Plunkett commented,

Det är allvarligt för likvärdigheten i betygsättningen att lärare kommer fram till så olika resultat (Isemo, 2011, para. 6).

2.3.2 Students

While it would seem teachers still have some way to go in reaching common ground on assessment, a parallel and related question is what students understand the assessment criteria to be. This raises pedagogical and didactical questions about how individual teachers can best convey the relevant criteria to their students.

As noted in the introduction, Cunningham (1998) views it as imperative that all criteria are explicitly defined for students before teaching is initiated, the argument being that without such explicit criteria students lack the necessary information, and the information that they have a right to, in order to perform at their best. This issue has also been in focus within higher education (Rust, Price, & O'Donovan, 2003). At first sight, this would seem to be an obvious point and a reasonable demand. The question is, what are the consequences of following this line of reasoning? Alison Wolf (2001) takes this up in the context of education standards in the UK. She argues that by simply stating the criteria and making them explicit does not necessarily clarify things. On the contrary, she maintains that in the development the system of competence-based assessment in the UK in any case, it has had the opposite effect. In the process of trying to define certain concepts and requirements, other criteria were also deemed necessary and subsequently introduced. According to Wolf, this has in practice led to an explosion in the number criteria and a system that is counterproductive to the purposes of both promoting student learning and simplifying assessment (Wolf, 2001).

Sweden has taken another approach. The attempt has not been made to specify each grade at the national level so that they define the content of teaching. That responsibility lies at the local level, primarily with schools and teachers.

A fundamental idea of the present curriculum is that the teachers in consultation with the students should decide on the teaching content. The idea is that the national goals and the criteria for the award of grades will become explicit for the students by being reformulated and incorporated into the format of the locally chosen teaching content. (Skolverket, 2005, p. 14)

Let us focus again on the goal to be fulfilled at the end of Grade 9 specifically relating to writing.

Eleven skall kunna begära och ge information i skrift samt berätta och beskriva något (Skolverket, 2000, "Mål som eleverna skall ha uppnått i slutet av det nionde skolåret" para. 5).

How does one recognize whether a text meets this requirement? Clearly, more information is required to make an informed judgement. It would seem however that any attempt to explicitly define it in terms of more detailed criteria is likely to be counterproductive, in light of Wolf's critique. Other studies also highlight the problems related to an overreliance on specified criteria (Kathryn Ecclestone, 2001; Rust et al., 2003) The theme that seems to repeat itself in the literature is that explicit criteria, while necessary, are in themselves insufficient to make the criteria transparent for students.

It would appear that other strategies are required. Fredriksen and Collins (1990) offer an alternative. Transparency for them is not solely reliant on definitions of criteria. Their paper, *A Systems Approach to Educational Testing* proposes a radical model for change at the national level in the USA. They advocate that assessment criteria "...must be transparent enough so that they [the students] can assess themselves and others with almost the same reliability as the actual test evaluators achieve (Fredriksen & Collins, 1990, p. 7). This transparency is, however, not reliant on how precisely the criteria are formulated. To enhance understanding, they also include methods for fostering improved performance, including self-assessment and feedback on performance. Another central feature of their proposal is the establishment of a library of exemplars of student work at a variety of levels, and each with a detailed rationale explaining why the specific grade is awarded. A training system for scoring different tests is also presented. 'The training materials can become the medium for communicating to teachers and students the critical traits to look for in good writing, good historical analysis, and good problem solving (Fredriksen & Collins, 1990, p. 5). Three groups need to learn the relevant criteria: The administrators, teachers, and students,"...who must internalize the criteria by which their work is being judged" (Fredriksen & Collins, 1990, p. 6).

The work of Fredriksen and Collins offers an alternative approach to clarify criteria – using such methods as self-assessment and exemplars. Their work is a precursor to the contemporary focus and research on formative assessment. Transparency, as outlined in their paper, is not a simple construct, but rather the result of a multifaceted approach to promote an understanding of the criteria and enhance learning.

Formative assessment is a pedagogical tool specifically aimed at generating feedback as a means of promoting student learning (Nicol & Macfarlane-Dick, 2006). In the book, *Handbook of Formative Assessment* Cizek defines it in the following terms:

Broadly conceived, formative assessment refers to the collaborative processes engaged in by educators and students for the purpose of understanding the students' learning and conceptual organization, identification of strengths, diagnosis of weaknesses, areas for improvement, and as a source of information that teachers can use in instructional planning and students can use in deepening their understandings and improving their achievement (Andrade & Cizek, 2006, p. 7).

In their important work on the subject, *Inside the Black Box*, Black and Wiliam (1998) examined 580 research articles related to formative assessment. They posed three questions:

- Is there evidence that improving formative assessment raises standards?
 - Is there evidence that there is room for improvement?
 - Is there evidence about how to improve formative assessment?
- (Black & Wiliam, 1998, p. 2)

After reviewing the literature, they answer ‘Yes’ on all points. This sparked renewed interest in the topic. Much of the recent research has been on the type and quality of feedback given (Price, Handley, Millar & O'Donovan, 2010; Nicol & Macfarlane-Dick, 2006). Feedback can take many forms, and utilize any number of mediums, including portfolios, rubrics, peer and self-assessments or exemplars (Skolverket, 2011c). However the binding element is that feedback, in whatever form, is always part of the social discourse. As Wiliam and Black (1998) put it, “We start from the self-evident proposition that teaching and learning must be interactive” (p. 1).

However, certain parameters need to be met for this to be effective. According to Sadler (1998) for feedback to be formative the students need to have “a) a concept of their learning goal; b) the ability to compare actual and desired performances; and 3) the ability to act in such a way as to close the gap” (Osmond, Merry & Callaghan, 2004, p. 274). Among other things, this process promotes the development of a meta-language to enable both students and teachers to define and compare the status quo and possible paths for development. It provides the parameters within which ideas can be tested, modified and applied –and not least, clarified.

What the literature seems to indicate is that formative assessment practices aid transparency. That is, they give a deeper understanding of what the criteria mean and how they can be applied. Formative assessment has shown itself to increase student outcomes in almost all learning situations (Sadler, 1998). Studies in the literature specifically focused on using formative assessment techniques to clarify criteria have also demonstrated its effectiveness (Jonsson, 2010; O'Donovan et al., 2004; Dochy, Segers & Sluijsmans, 1999).

The social constructivist orientation of formative assessment finds theoretical support in Vygotsky, and in particular his ZPD (Zone of Proximal Development).

When we talk about working in the zone of proximal development, we look at the way that a child's performance is mediated socially, that is, how shared understanding or intersubjectivity has been achieved. This includes the means by which the educator reaches and meets the level of the child's understanding and then leads the child from there to a higher, culturally mediated level of development. (Verenikina, 2003, p. 5).

For Vygotsky, learning takes place within a certain zone for each individual. To promote learning the educator needs to first establish what the person already knows so as to be able to guide him or her to a deeper understanding. Here we see that the 'social dialogue' is fundamental the learning process.

Exactly how teachers promote the clarification of the assessment criteria and grades at the local level is not the focus of this study. I have pointed to two theoretical approaches here and note that a model using formative assessment techniques appear to offer advantages over attempting to define the criteria in detail. The research shows that the latter technique can promote transparency by providing the social context for developing a common understanding of what the criteria mean in practice. After Skolverket's report in 2000 noted above, Skolverket published Allmänna råd 2004 to support teachers and promote fair and standardized assessment. A number of other publications related to the national tests are also now available, including Diagnostic Material and individual exercises and tests (Skolverket,

2011). The question is, how far have we come in terms of establishing common ground in assessment for students and teachers? To what extent are the criteria transparent?

A recent study by Helena Tsagalidis (2008), *Yrkeskunnandets kinesiska ask* examines the extent to which common assessment criteria are used in the context of secondary school's Hotell- och restaurang-program (HR-program). She looks at the responses of both teachers and students within categories elicited from interview data. She also documents how these responses relate to the grades G, VG and MVG. In concluding, she notes that students and teachers' views differ in many respects about what is deemed to be important within each category. Perhaps the most worrying aspect however, is that a number of the categories elicited from the data cannot even be related to the subject's course plan or national goals (Tsagalidis, 2008, p. 152).

This study, Common Ground, aims at contributing to the debate on transparency. Paradoxically, I attempt to examine this by giving the students texts they have never seen and then compare the criteria they use to assess these texts with a list of text analysis criteria from Skolverket that they have never seen either. The idea is that the internalized criteria that they arrive at are the product of the 'process of clarification' that has taken place in the classroom.

3 Method

This section outlines the instruments, procedures and methodologies involved in collecting and analyzing the data for this study.

3.1 Data Collection and procedure

It is vital that the process of data collection is carefully thought out and controlled so as to minimize unplanned influences and ensure that one measures what one has set out to measure. Below is a brief overview of the different elements involved in data collection for this study. It includes the target group of the study, the criteria employed in the exercise, the exercise itself as well as the process of its development, the role of focus groups in the process and lastly, how the exercise unfolded in the classroom.

3.1.1 Focus Groups

Which methodological approach would best suit the nature of the task at hand and fulfil the goals that have been set? This is not an easy question to answer given that the task is text analysis and assessment. In order to maximise student responses, focus groups are used as the basic model for data collection.

In the context of my study, focus groups offer a number of advantages over other methods. Firstly, they actively promote communication and as such enable participants to clarify their own thoughts on issues that may not come to light in a one-on-one interview situation for instance. They also promote a more relaxed environment for the collection of information than many other forms. Peer groups are able to sit undisturbed and discuss topics in an informal setting which can promote a freer flow of information both between participants and from the groups (Kitzinger, 1995).

Another positive feature is the democratic nature of focus groups. The emphasis is on oral communication within the group and as such those who don't feel confident providing written

responses can be an active group participant in any case and contribute in ways that may not otherwise be realised using more conventional methods (Halkier, 2010). There are drawbacks too, however. Group dynamics may not always be positive in these situations making debate and the free-flow of ideas difficult to achieve. In some cases, people may even be socially intimidated and effectively silenced. These situations can be mitigated if the groups are carefully selected however group dynamics cannot be totally foreseen or controlled (Kitzinger, 1995).

In weighing up the potential positives and negatives regarding the central task of the study, I chose to use focus groups as they seem to offer the best chance of maximising the discussion and feedback relating to the texts in the study.

3.1.2 Participants

Although my target group is actually all Grade 9 students in Swedish secondary schools, I am naturally forced to select only a small number of them due to resource and time limitations. My selection is primarily driven by my social network and physical location. I have chosen to work with students from three schools. The first is the one where I have worked for the past eight years. I selected two grade nine classes, one of twenty-two students and the other of sixteen students. The second school is in the same town but has fewer pupils in Grade 9. There, I did my study with a class of six students. The third school is in another municipality, where I had access to three grade 9 classes each of which had between eighteen to twenty-two students. All three are independent schools. This was not a conscious factor in my selection, but rather an indirect consequence of the network I have built up as a result of my work experience in Sweden.

On the specific days that I carried out my study at the different schools, not all students from the respective classes were present. This was due to a variety of reasons, including illness, visits to prospective secondary schools as well as failure to return the letter signed by the legal guardian which authorised participation for those under fifteen. Total losses were 11 students from a total of 106, which equates to ten percent. In total, ninety-five students took part in the study.

3.1.3 Assessment criteria

The Teacher's Booklet "Bedömning och exempel" provides fourteen example-texts which have all been analysed, commented on, and assessed by

... en erfaren grupp bedömare utifrån kursplanens mål, nationella kriterier för betygen Väl godkänt och Mycket väl godkänt, inklusive allmänna råd för bedömnings inriktning, samt övriga faktorer som anges ovan. (Skolverket, 2008, p. 21)

These "other factors" relate to what Skolverket calls "bedömningsfaktorer" (Skolverket, 2008, p. 21). They are grouped into two categories; Content and Language.

Innehåll

1. Om texten ger uttryck för en vilja att använda språket för att förmedla ett innehåll
2. Om texten är sammanhängande och strukturerad
3. Om innehållet är fylligt eller magert/torftigt
4. Om ämnet är utförligt behandlat - om eleven fokuserar eller bara ytligt behandlar ämnet/de olika punkterna
5. Om texten är anpassad till mottagaren/syftet

Språk

6. Begriplighet – förmåga att uttrycka ett budskap klart och tydligt
 7. Ledighet, variation och säkerhet – flyt
 8. Strategier att ta sig runt språkliga problem
 9. Vokabulär och idiomatik (omfång, variation)
 10. Meningsbyggnad – förmåga att binda samman satser och meningar
 11. Korrekthet (vokabulär, idiomatik, grammatik och stavning)
- (Skolverket, 2008, p. 21)

What Skolverket means is that texts can be analysed and graded using ‘bedömningsfaktorer’, in conjunction with the other documents relating to assessment mentioned above. That is exactly what Skolverket’s ‘experienced group of assessors’ has done. Each text has been assessed on the basis of the goals of the course plan, the national criteria for VG and MVG, the advisory document relating to assessment, as well as the ‘bedömningsfaktorer’ laid out above. As such, the comments, analysis and assessment that Skolverket provides for each text can be seen as a concrete expression of the relevant assessment criteria for each of the texts. Thus, Skolverket’s assessment criteria for each text are manifest in the comments (or responses) provided by the ‘experienced group of assessors’.

Similarly, student responses to the texts are deemed to embody their own assessment criteria. I utilise the ‘bedömningsfaktorer’ above to order the responses from both Skolverket and the students. In doing so, ‘bedömningsfaktorer’ provide a benchmark for a common set of assessment criteria for both Students and the ‘experienced group of assessors’ from Skolverket. They provide a system for mapping student responses to the texts and establishing which elements are viewed as significant. In this context, the student responses represent a concrete and detailed expression of what their assessment and grading is based on.

Thus, even though the ‘bedömningsfaktorer’ simply classify the responses of both students and Skolverket, for the purposes of this study, they define a *common set of assessment criteria*. That which constitutes a G, VG or MVG text is defined via these criteria. For both the students and Skolverket, the responses to the texts constitute a concrete expression of their own assessment criteria. Student responses are based on their ‘internalised criteria’, whereas Skolverket’s responses are formulated from all relevant official information related to assessment. The ‘bedömningsfaktorer’, therefore simply group the existing assessment criteria provided. For this reason, hereafter I will use the term *assessment criteria* to relate to ‘bedömningsfaktorer’ (1-11) and *category* to relate to the two groups of criteria: Content (1-5) and Language (6-11).

I have chosen to use these two categories and the eleven related assessment criteria as the basis for my analysis for the following reasons. Firstly, they provide a broad platform for text analysis. Secondly, they have been used by Skolverket’s assessors and have therefore already played a role in the analysis of the texts and related comments. Further, as I intend to analyse Skolverket’s comments as well as the students’, there is an added advantage in using the same system of classification – it ought to make for an easier fit between the comments and the criteria. Lastly, and as noted earlier, it makes it possible develop a common set of assessment criteria.

Along with the assessment criteria, this study also uses the fourteen texts in the teacher’s booklet as the material for student analysis. The fourteen examples of student texts range from G to MVG. Positive and negative signifiers are used to further define levels within each

grade. As such, there are in practice nine grade-levels but only three grades. Under each text there is a written analysis and a brief summary. Lastly, each text is given a grade-level.

3.1.4 Developing the Exercise

Firstly, I examined Skolverket's material and organized the analyses, comments and grading of the texts in a new way. I went through all the comments which are written in free-flowing text and broke down the sentences into points that focused on strengths or weaknesses. I then classified these elements under what I considered to be the most appropriate assessment criteria out of the eleven mentioned in the section above. After that, I included the text summary which is found at the end of each of the assessments. Finally, I noted the grade allocated to each of the texts. (See appendix 1 for an example of the exercise.)

The aim was to formulate a structure which could be used for comparative purposes. The students were going to be given a difficult task. It would be unrealistic for me to expect them to evaluate, assess and grade texts without any explicit guidelines or criteria. The objective was to formulate a very basic structure which marries in with the information provided by Skolverket but doesn't overly influence or direct what the students themselves have to say about the texts. The conscious decision was made not to include any explicit assessment criteria at all in the student exercise as one of the main aims of the study is to elicit the students' own criteria.

Firstly, a pilot study was done to assess the strengths and weaknesses of the proposed exercise. As a result, the following conclusions were drawn. It would be beneficial to:

- Modify the layout of the template in order to make it clearer.
- Increase the time allocated to complete the exercise to ensure that the students were to get the full benefit of comparing their assessments to Skolverket's.
- Simplify instructions.
- Spend more time on eliciting criteria from students before starting the group work.

Subsequently, the necessary modifications were made and the principals of the schools were contacted to organize times in order to implement the study. I was fortunate enough to be invited to do my study at all three schools.

3.1.5 Procedure

Prior to going to class I asked the teacher to help me group the students – in groups of two to four – in such a way as to promote communication within the group. After my introduction, I asked the class the following questions; “What is it about a text that makes it a good text? What sort of qualities might a good text have?” My aim was to elicit as many responses as possible so as to give all groups the benefit of the class's responses and hopefully generate ideas for discussion later on. Then I displayed some instructions relating to the group work and the texts and went through them verbally.

The process was as follows. In their groups the students decided on who should record the group's evaluations and assessment. The students were then informed about the structure of this part of the national test and how the questions are formulated, after which, they got a copy of Part C of the 2008 test and got the chance to ask questions about it. Then, each group was given a folder with four copies of a text as well as the template for their text analysis and assessment. Each group was then asked to read their text individually then discuss it focusing

on its strengths and weaknesses. After that, one member recorded, in point form, the positive and negative aspects of the text discussed. A final summary was then requested and lastly, a grade for the text was to be provided. If consensus on a common grade could not be reached within the group, individual grades were to be recorded.

When groups had completed the first text, the folders were collected and the second text was distributed, and the same process was repeated with the new text. When the exercises for both texts had been completed I then handed out Skolverket's assessment and grading of the same texts and we discussed the similarities and differences relating to the different assessments.

Part C has two questions. There are fourteen student example-texts, seven relating to each question. All groups received example-texts relating to the same question in Part C at the same time. This meant that each session had a maximum of seven groups. The sizes of the groups depended on the number of participants, but ranged from two to four. Each group received different texts to work on. These texts were randomly distributed to the different groups.

3.2 Data Analysis

My method of analysis includes both quantitative and qualitative dimensions. The process of classifying the comments for each example-text within the parameters of the template for my exercise was inherently qualitative. I was required to break down the free-flowing text into distinct elements which, in my view, best matched the assessment criteria in my template. Once the exercise was complete and the data had been collected from the students, I then grouped their responses as well according to the same criteria. Using an Excel document, I was then able to use quantitative methods to measure the number of responses within the different categories and compare the statistical information.

3.3 Reliability, Validity and Generalizability

The concept of reliability in the context of academic studies relates to the extent to which the procedures employed resist the inclusion of irregularities in the data. I have sought to achieve a high degree of reliability by adhering to the following procedures. Firstly, I attempted to minimize the inherent interpretive element of this study by employing Skolverket's material as my benchmark. Secondly, I developed my exercise in a way that incorporated the 'bedömningsfaktorer' already employed by Skolverket. Further, I implemented a pilot of my exercise to test its appropriateness and effectiveness in fulfilling the goals of the study, following which I modified certain aspects to improve the exercise's effectiveness.

Despite my attempts, certain subjective elements remain which do bring into question reliability. As my study involves assessment it inevitably also involves interpretation. Interpretation, by definition, implies variance. The most difficult aspect of this study for me as the collator of the data was to interpret and classify it correctly. Firstly, I was required to interpret and classify the comments provided by Skolverket into their 'bedömningsfaktorer'. Then, when the students had completed the exercise, I was required to do the same thing again for each group's responses. This was no easy task and without question involves an interpretive element that raises the issue of reliability. More credibility could be attributed to this aspect of the study if it were done by more than one person, either individually or as a group.

Another factor which could be seen as weakening the reliability of this study is the fact that I am the English teacher for two of the classes that have participated in this study. In response, I can only say that I was conscious of this potential conflict of interest and did my utmost to ensure that I didn't prejudice the study in any way.

That which lends weight to the reliability of the study is my experience as a teacher. I have many years' experience of working with the national tests and related criteria (Patel & Davidson, 2003). In addition, I developed a system of classification using an Excel document that enabled me to maintain a high level of control of the data. Using this program helped me minimize inconsistencies in classification relating to the assessment criteria.

The notion of validity raises another issue. Here the question is whether the study measures the thing that it is claimed to measure (Patel & Davidson, 2003). The intention of the exercise was to elicit how the students themselves assess texts, and to document the central criteria they use. To this end, the students received no extra input other than listening to each other's thoughts about what constitutes a good text prior to their focus group discussions.

I would maintain that the results express a high level of validity. The template for the exercise itself offered no other guidance than breaking down the comments into positive and negative categories. As such, the students were required to assess the texts using their own criteria and language, with nothing else imposed or prescribed.

The third fundamental issue to evaluate is whether or not the conclusions drawn have relevance beyond the parameters of the specific study in question. Do they have a more general applicability and if so to what extent (Patel & Davidson, 2003)? As noted earlier, ninety-five Grade 9 students from three different schools took part in this study. While this is not a large number it is perhaps sufficient to suppose that some of the central themes in the findings may have a broader relevance. I would also suggest that these findings may also have relevance for the Lgr11.

3.4 Ethical Issues

Firstly I checked the internet to ensure that the test I intended to use was not under any confidentiality restrictions. To double-check, I also contacted Gothenburg University and spoke with the responsible personnel regarding my study. I was assured that I was able to use the test for my purposes. My next step was to contact the schools I had in mind.

After gaining permission to carry out the study at the three schools, I forwarded a letter of introduction as well as a letter for the legal guardian of those students who had not yet turned 15 (See attachment 2). I clearly explained to both teachers and students that the study is voluntary and that not only the individual schools but also individuals that participate will have anonymity. I further explained that all related information would be handled confidentially and according to the appropriate ethical regulations. No names were collected or recorded. The information collected can in no way be traced to the individuals or groups that participated. These measures that have been followed ensure that the four central demands relating to academic inquiry have been met (Vetenskapsrådet, 2009).

4 Findings

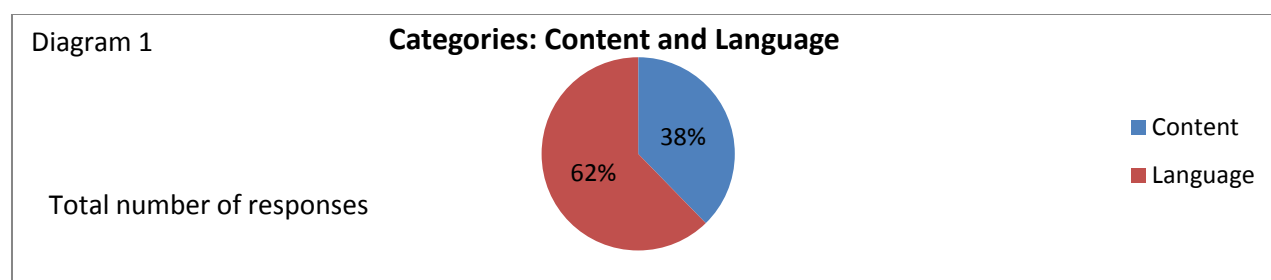
Here, I document the results for each of the research questions in turn and then move on to analysing the data in the following section.

4.1 Student Responses

The results in Section 4.1 relate to student responses. There are three points of focus. The first is on the two major categories, the second on the assessment criteria and the third is on the positive and negative responses for each criterion.

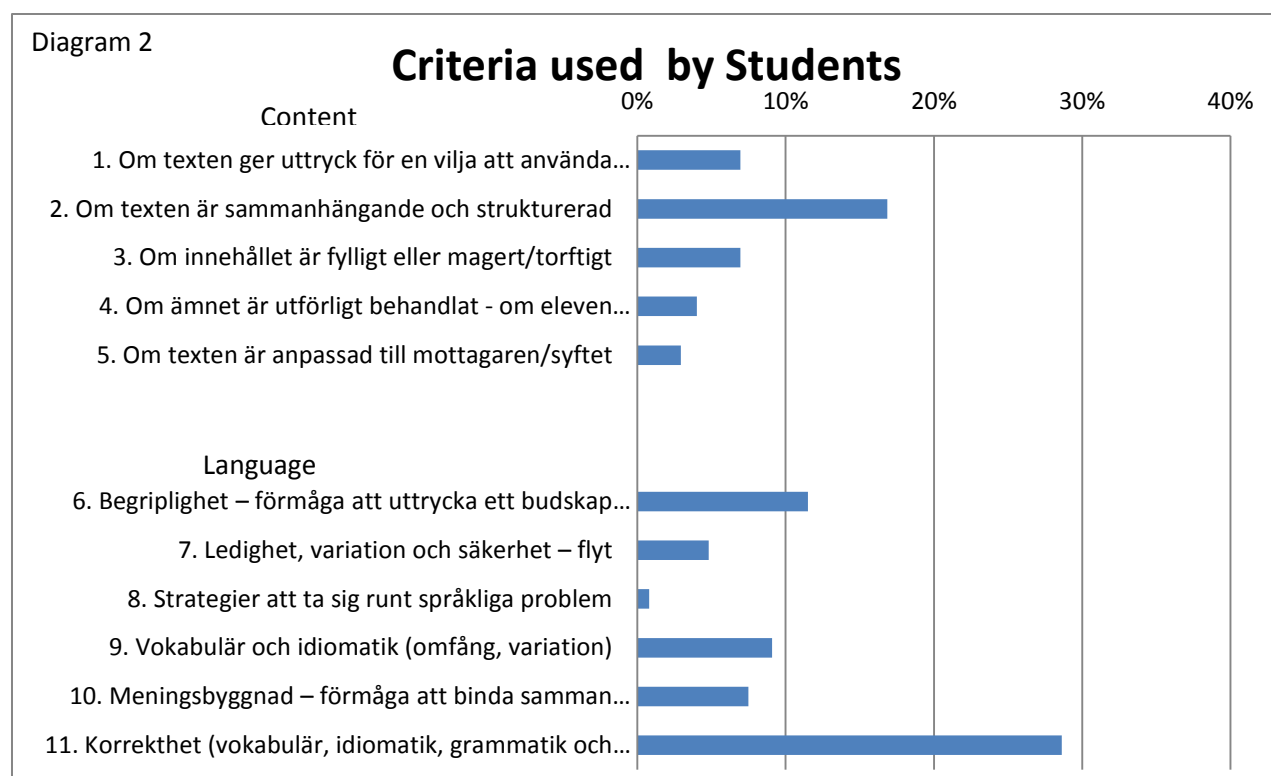
4.1.1 Categories: Content and Language

Diagram 1 below shows the number of responses recorded in each of the categories, content and language. This clearly shows that student responses are heavily weighted in favour of language over content.



4.1.2 Assessment criteria used by Students

Diagram 2 displays the number of responses for each of the assessment criteria. The total number of responses is 374, of which 141 relate to content and 233 relate to language. This demonstrates that the vast majority of student responses related to Language. Just 38% were focused on Content.

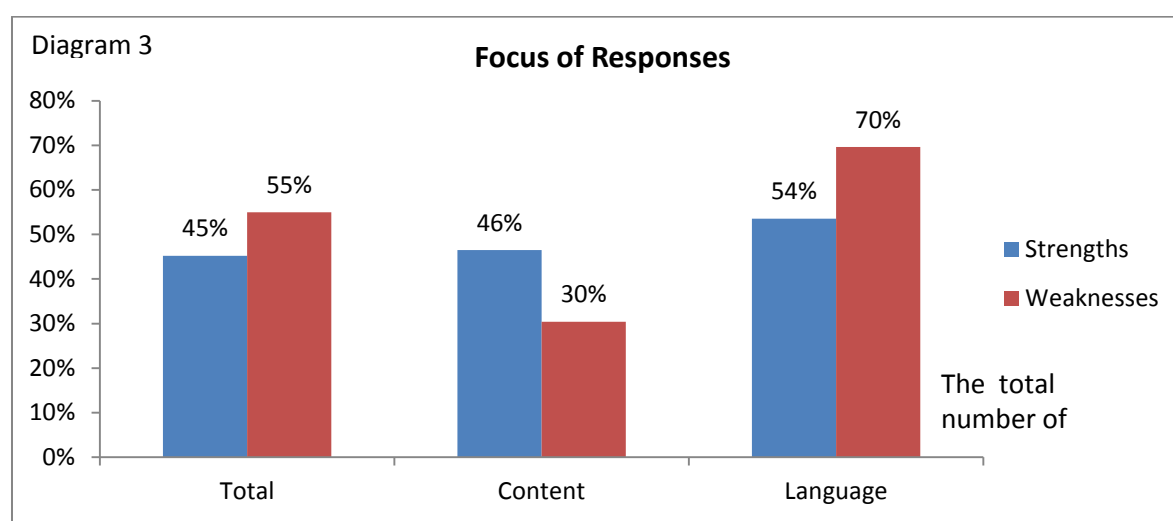


The Content category is dominated by responses relating to the second criterion, which relates to a text's coherence and structure. Over half of all content-related responses, 73 of 141, fall into this group. This figure translates to 17 percent of the total. The other responses are fairly evenly distributed within the category. Distribution is between three and seven percent of the total number of responses.

Results for the Language category show an even wider divergence. Two assessment criteria have responses that are higher than ten percent of all replies. *Comprehensibility-the ability to clearly express a message* receives 43 responses which amount to 11 percent of the total. The other criterion, relating to Correctness, has by far the most responses of all the criteria in both categories. This criterion received 107 responses, or 29 percent of the total number recorded. Another significant point is the relatively few responses recorded under criterion 8, which relates to *strategies to get around language problems*. Three responses are recorded here, which amounts to just one percent of the total.

4.1.3 Focus on Strengths and Weaknesses: A Comparison

Diagram 3 shows the percentage of responses that have a positive respective negative focus within each category, as well as the total percentage figures for all student responses.



As a whole, the students have provided more responses that focus on weaknesses than strengths. However, when we look at each category in isolation we see that the results diverge a great deal. Students have responded positively in terms of Content but negatively in terms of Language. The higher percentages for Language reflect the larger numbers of replies in that category (as noted previously in Diagram 1), however, these quantities have nothing to do with the positive and negative focus within each category. The figures indicate that, overall, the volume of comments with negative focus outweighs the positive and that the Language category includes by far the largest number of the negatively orientated responses. However, students have recorded much higher numbers of positive responses than negative relating to the Content category.

4.2 Categories: A comparison

The results in this section show student responses in comparison to Skolverket's responses. As in the previous section, there are three points of focus. The first is on the two major

categories, the second on the assessment criteria and the third is on the positive and negative responses for each criterion.

4.1.3 Categories

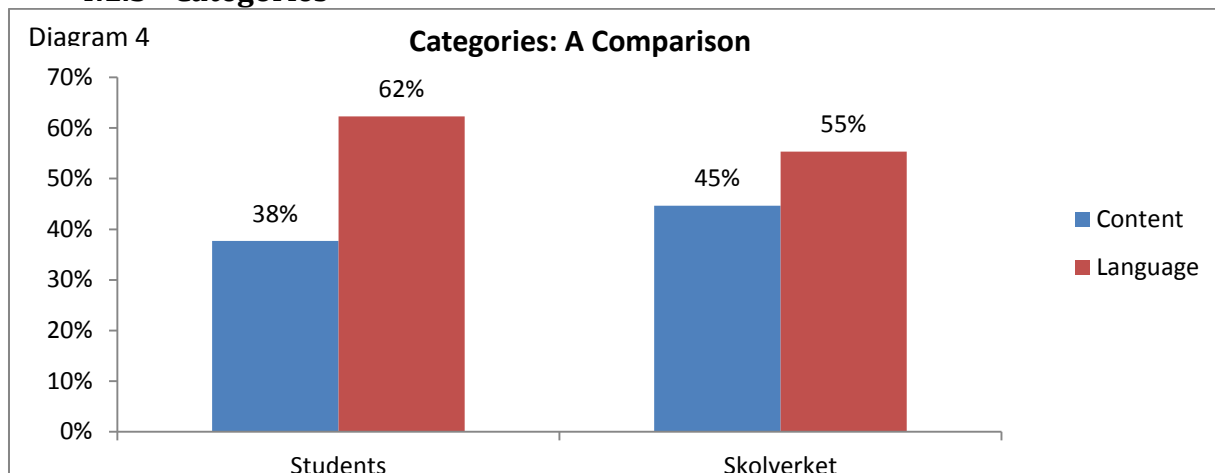


Diagram 4 shows that the dominance of the Language category regarding student responses is reflected in Skolverket's responses as well, but at a reduced level. That is to say, Skolverket has proportionately more replies relating to Content than students do, however the Language category attracts the most responses for both students and Skolverket.

4.1.4 Assessment criteria

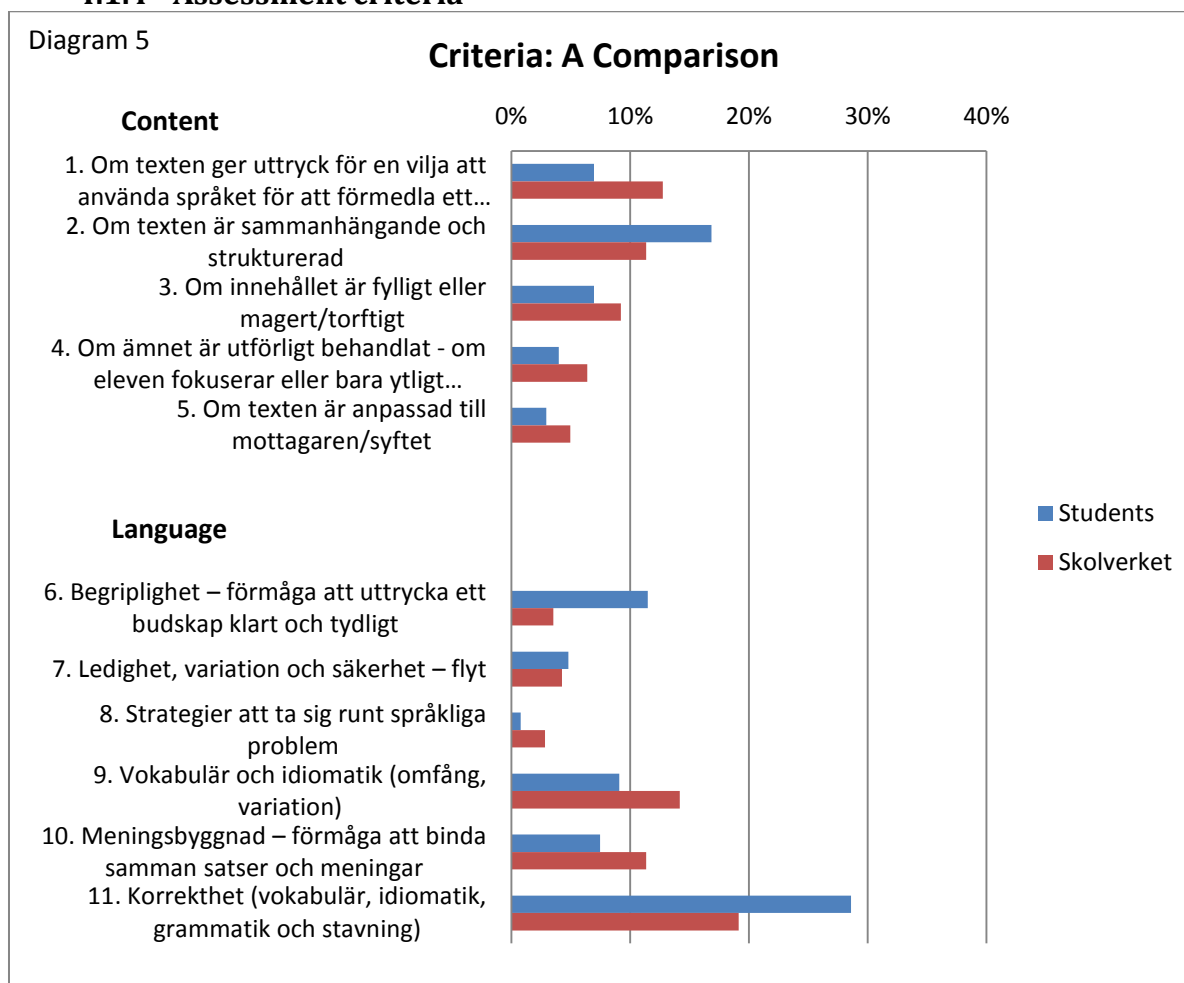
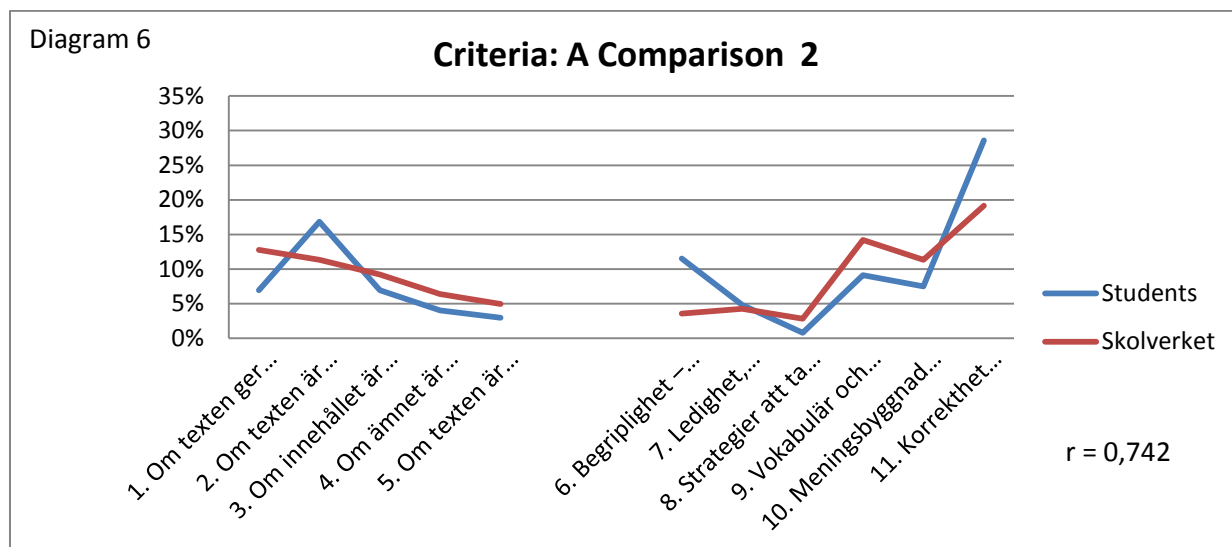


Diagram 5 shows that there is a general uniformity in the responses from students and those given by Skolverket. That is, criteria that receive for example a high number of responses from the students also receive a proportionately high number of responses from Skolverket. This relationship is illustrated in the line graph below (Diagram 6). This shows that there is a strong correlation ($r = 0,742$) between the responses of the two groups.



While a strong correlation is demonstrated, there are specific points of divergence that can also be seen. The responses for text one, two, six and eleven show most divergence from the general trend. Text one relates to whether the text demonstrates a desire to express something, two relates to structure and coherence, text six relates to comprehensibility – the ability to clearly express a message, and text eleven refers to correctness in terms of vocabulary, idiomatic expressions, grammar and spelling. In the first case, the percentage of student responses is proportionately lower, but in the three other cases, the percentage of student responses is proportionately higher than those from Skolverket. Because student responses have this focus, the students also display lower percentage points than Skolverket in many other criteria.

4.1.5 Focus on Strengths and Weaknesses

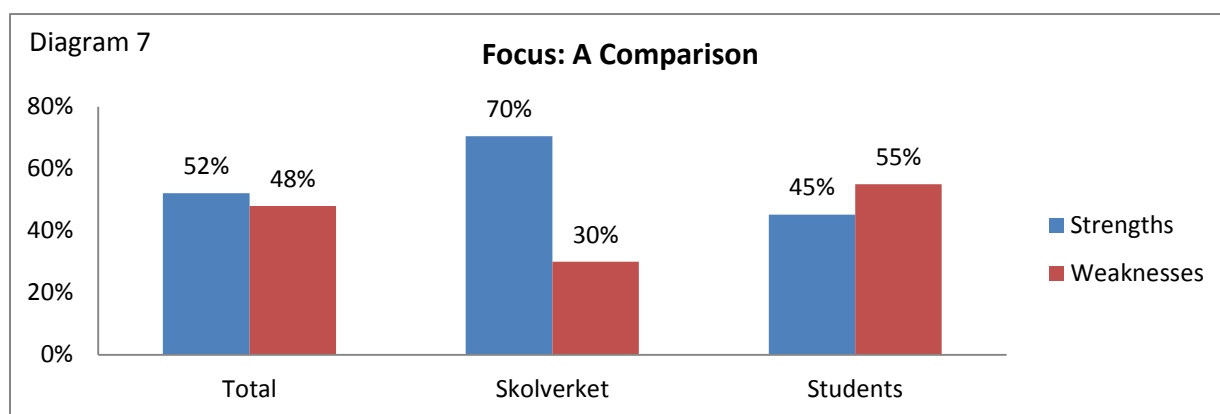


Diagram 7 shows that while the total number of responses is fairly equally balanced, student responses are much more negatively orientated than those from Skolverket. Over half of the student responses are negative (55%), whereas only 30% of Skolverket's have that

orientation. In other words, 70% of Skolverket's responses, but only 45% of student responses, are positive in focus.

4.1.6 Focus on Strengths and Weaknesses within each Criterion

Diagrams 8 and 9 break down the information relating to strength and weakness orientation even further by examining the data in terms of each category and criterion.

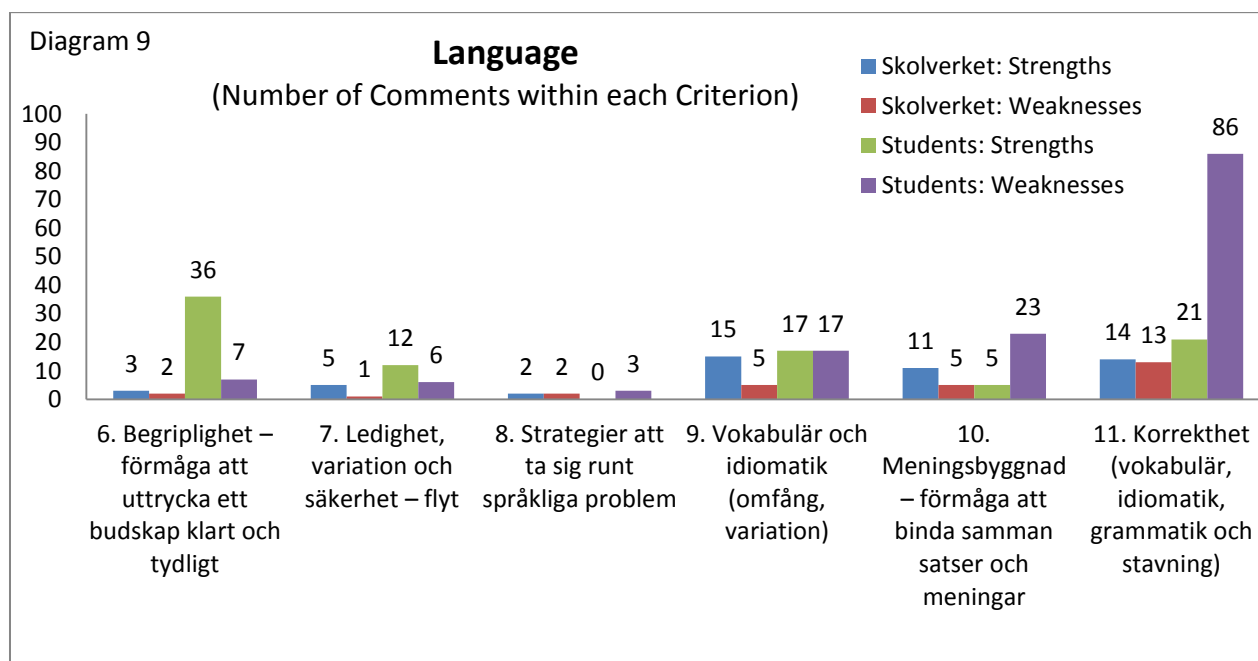
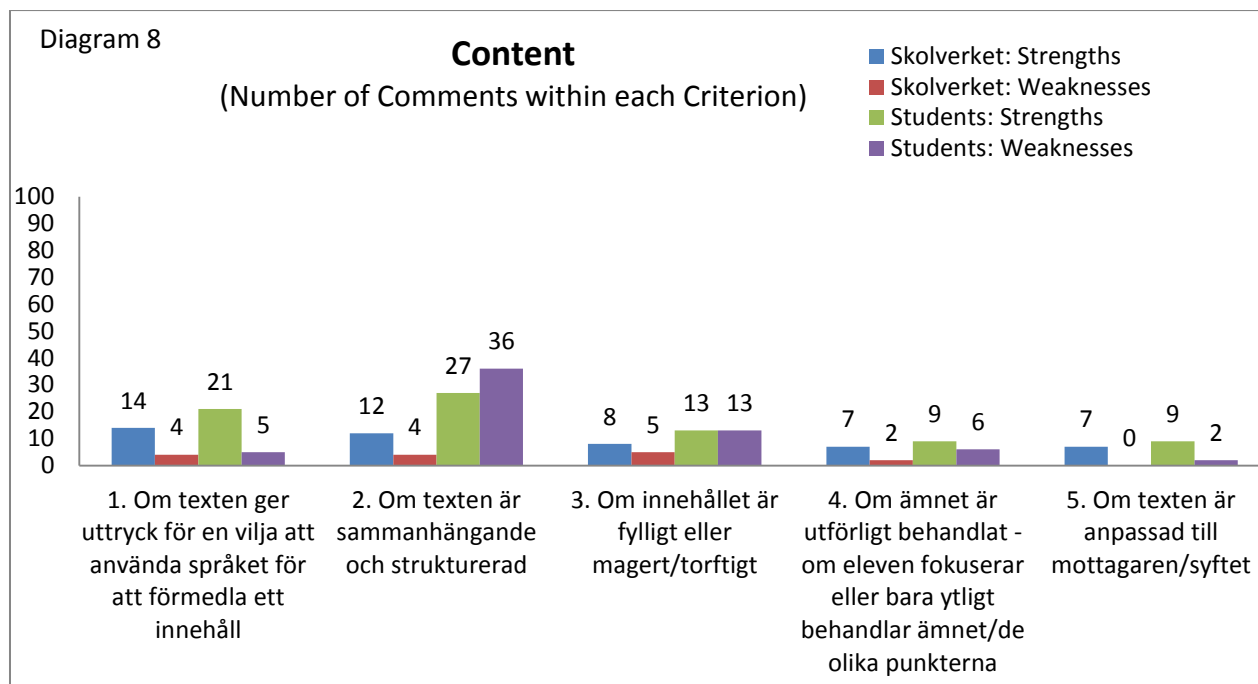


Diagram 8 and 9 illustrate that Skolverket's responses are positively weighted for each criterion, with one exception – text 8 – which has only four responses and is equally divided between the positive and negative. As noted earlier, student responses are predominately negative in focus. What these diagrams also show is that the spread of positively and

negatively orientated responses is not uniform for the students. Some assessment criteria are viewed very positively while others have a dominant negative focus. When we examine the three criteria which are overrepresented compared to Skolverket's responses (criterion 2, 6 and 11) we see that criterion 2 is a little negatively weighted in terms of responses and criterion 11 very negatively weighted. Criterion 6, on the other hand is very positively weighted.

4.2 Grades: A Comparison

This section begins with an overview of all grades set by students and Skolverket. Then the data is presented using a number of different parameters which compares student grading to Skolverket's recommendations.

The example-texts are grouped according to the two questions on the national test, where the students get to choose one of the two options. . Diagram 10 relates to example texts 1-7, all of which constitute responses to Question1: *In a World Full of Things*. Diagram 11 relates to example-texts 8-9 and Question 2: *Proud Of....* The first bar of each text represents Skolverket's grade for that specific text. The other bars represent grades set by students. Some texts have more bars than others because in cases where the groups could not reach agreement, individual grades were set by each student in the group.

4.2.1 Overview

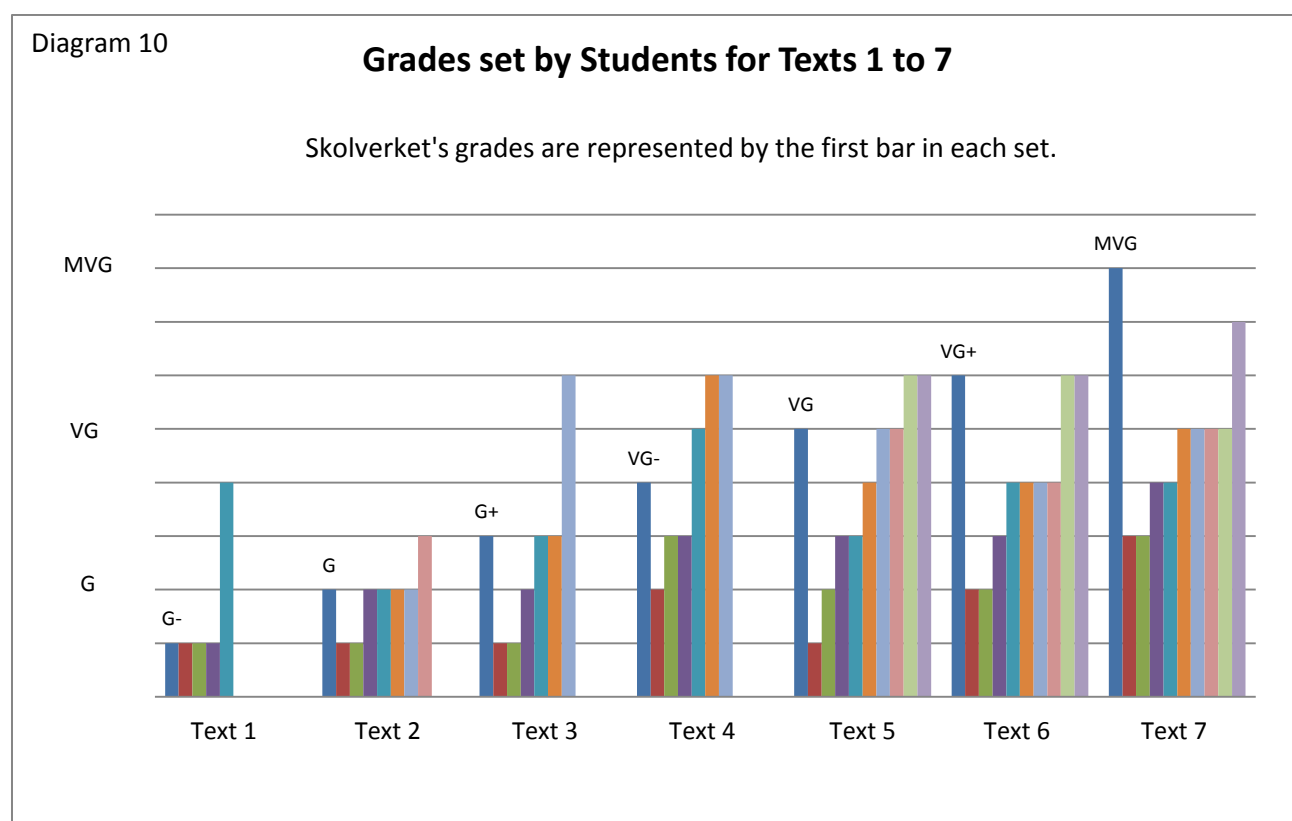
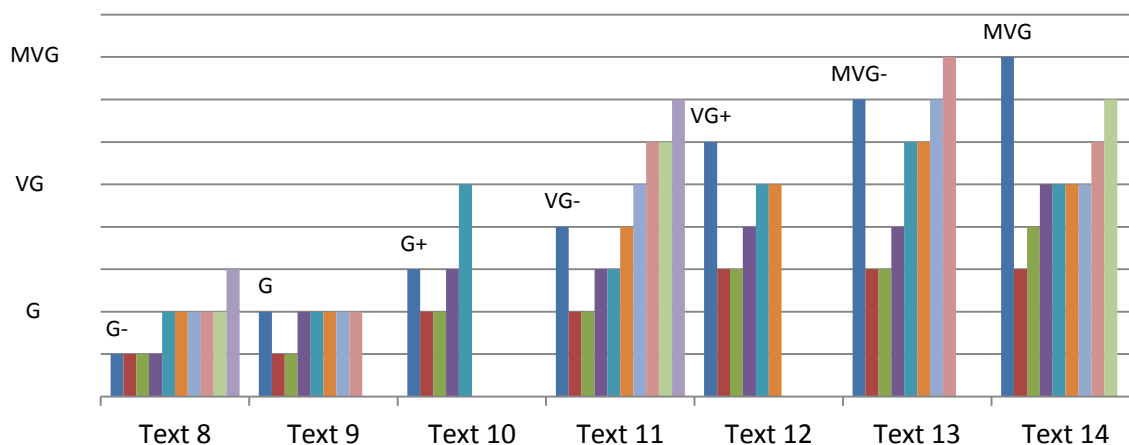


Diagram 11

Grades set by Students for Texts 8 to 14

Skolverket's grades are represented by the first bar in each set.



Ninety-nine grades have been recorded relating to the 14 texts. Skolverket's assessment is that six of the texts fulfil the goals for Godkänt, five for Vål godkänt and three for Mycket väl godkänt. In the following section, *Various frames of reference for Comparison* the data from these diagrams is represented with specific points of reference.

4.2.2 Various Frames of Reference for Comparison

Diagram 12 shows the total percentage of grades that are exactly the same as Skolverket's recommendations using a nine-point grading scale. It also gives breakdown of this total in terms of the three whole grades of Godkänt, Vål godkänt and Mycket väl godkänt. As can be seen, approximately a quarter of all student grades matched Skolverket's exactly. If we examine perfect matches in terms of whole grades, we find that forty-nine percent of all Godkänt grades (G-, G, G+) set by students were exact matches but only eight percent of VG (VG-, VG, VG+) and four percent of MVG (MVG-, MVG, MVG+)

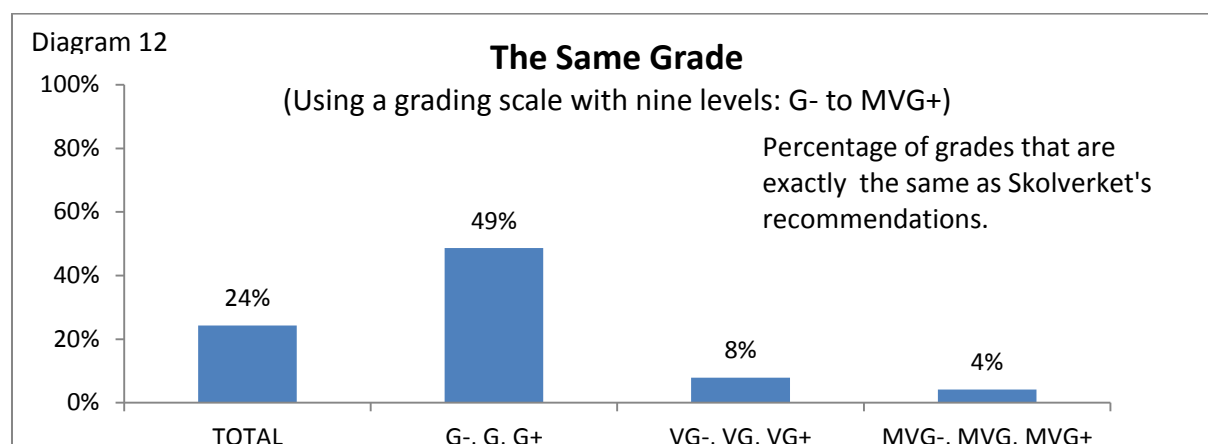


Diagram 13 shows the same thing as diagram 12, but with an inbuilt tolerance factor of one third of a grade-level. That is to say, these figures include student grades that are up to one

third of a grade higher or lower than the grade allocated by Skolverket. Using these parameters we see a marked increase in accuracy, with over half of all grades meeting this criterion and a full eighty-four percent of all grades at Godkänt level (G-, G, G+). Accuracy falls markedly at VG level and significantly at MVG level.

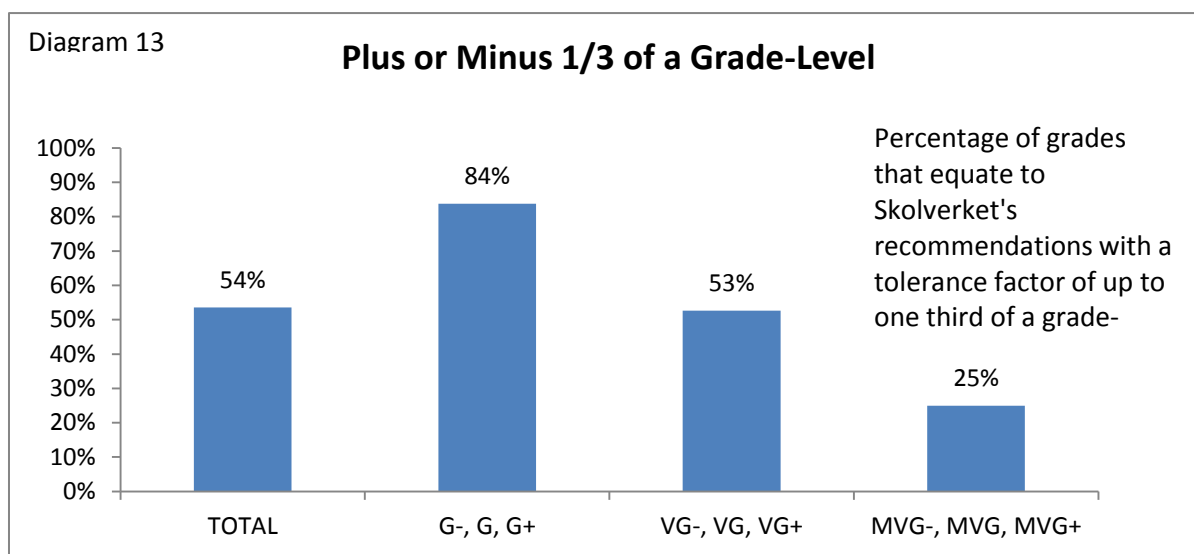
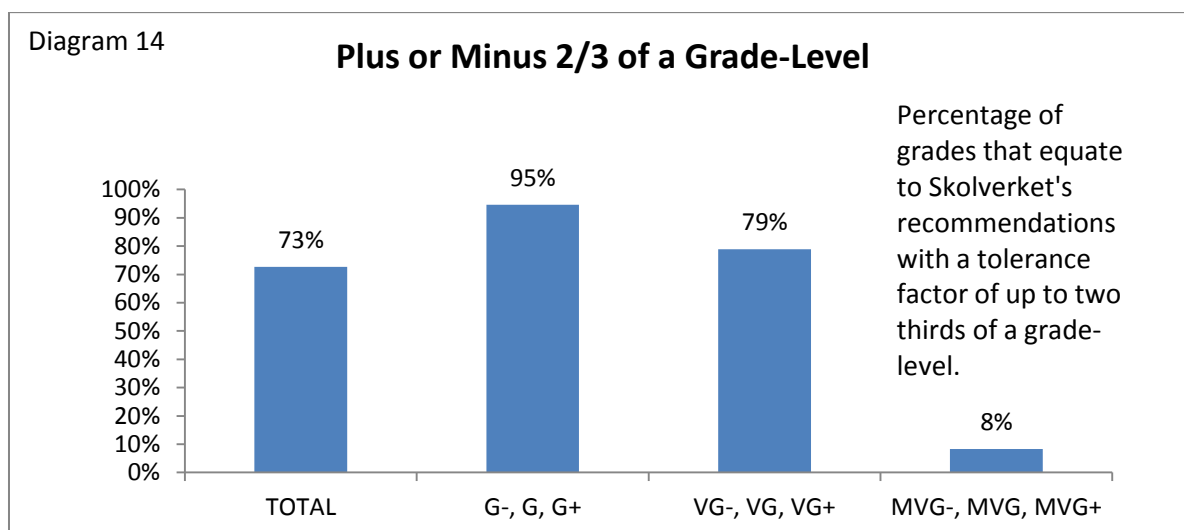
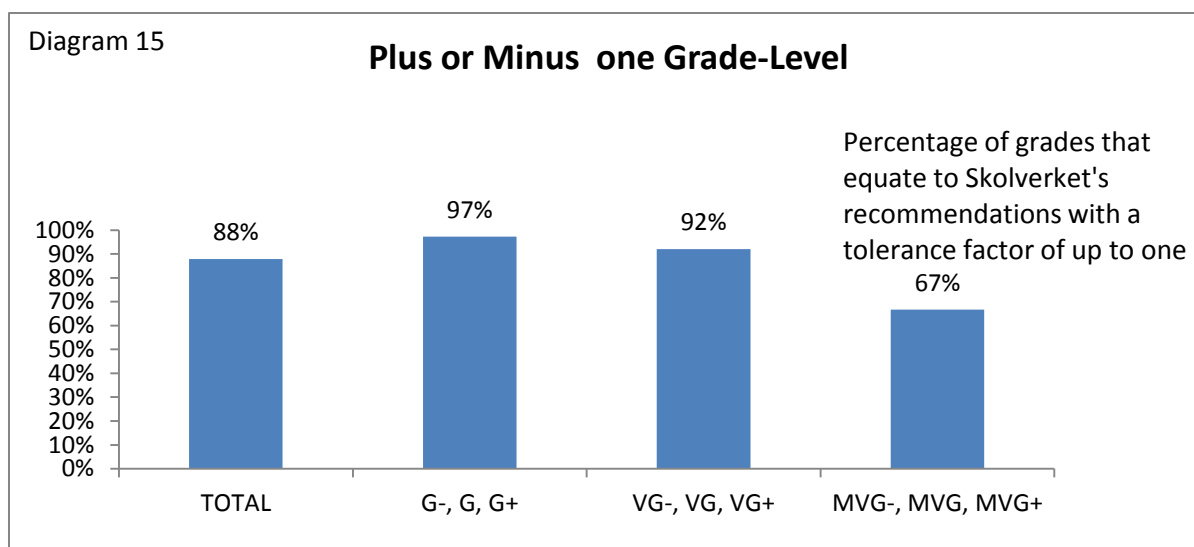


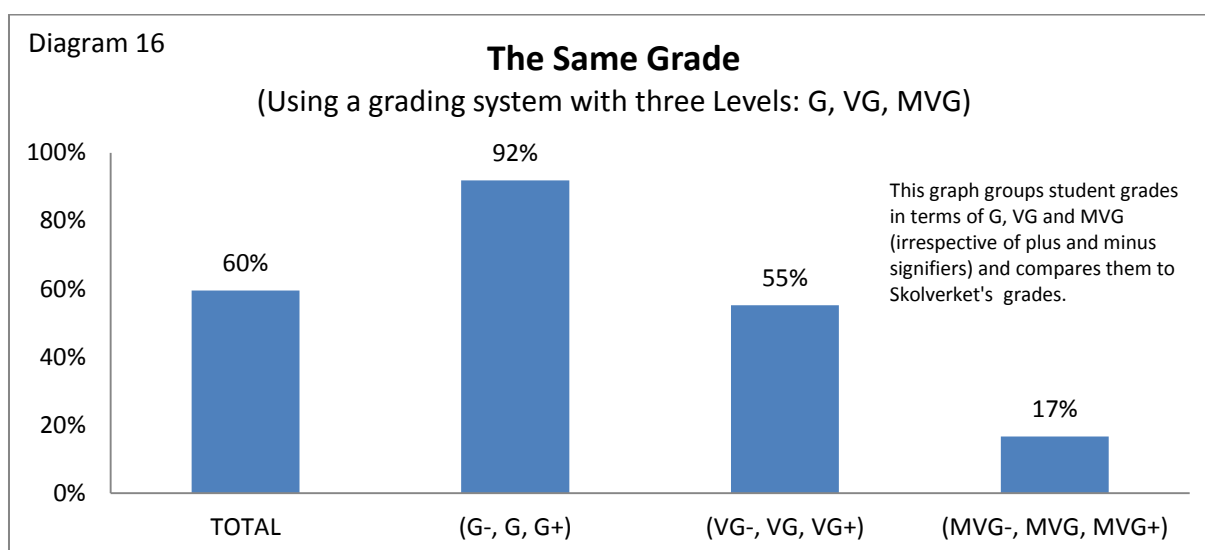
Diagram 14 extends the same principle to two-thirds of a grade level. Accuracy levels are again significantly improved using these broader parameters. Here we see that almost three quarters of all grades meet this criterion and a full 95 percent of all grades at the Godkänt level. As previously, accuracy drops off rapidly for the higher grades.



With diagram15 we see the result of the same correlation but with a tolerance factor of one full grade level. Here we can see that eighty-eight percent of all grades set by students are not more than one full grade-level from that advocated by Skolverket. We can also see that the percentages are high for the breakdown grades as well, and significantly, the pattern that we have seen in the data previously holds true here as well; the higher the grade the greater the inaccuracy of student grading in as measured against those advocated by Skolverket.



This final diagram of this series shows student accuracy using only three grades (G, VG and MVG). In other words, this diagram shows the accuracy of real grades. If a text is given VG- then the grade given is VG. The same applies to VG+. As such, plus and minus signifiers play no role in the final grading. Using these parameters we see that sixty percent of all texts match Skolverket's grades perfectly with a ninety-two percent fit at the Godkänt level. The inverse gradient relating to the higher grades is clearly displayed again with this data.



5 Analysis

In this section I examine each research question in turn and then attempt to synthesize the findings.

5.1 Common Assessment Criteria

Text analysis involves the use of a metacognitive language to classify the data (Shepard, 2000). Rather than elicit an independent set of assessment criteria from the data I have chosen to utilize Skolverket's 'bedömningsfaktorer' as the benchmark for reasons outlined in the Method section. Of necessity, these 'bedömningsfaktorer' are very broad so as to be inclusive.

Nevertheless, they do define specific aspects of a text that are deemed to be of central significance.

It is perhaps to be expected that Skolverket's responses are focused on and utilize the assessment criteria used in this study. After all the "expert group", which had the task of analyzing, commenting on and assessing these texts, had open access to the 'bedömningsfaktorer' on which to base their assessments. Furthermore, as noted earlier, they not only had open access to the assessment criteria used here ('bedömningsfaktorer'), but also to:

...kursplanens mål, nationella kriterier för betygen Väl godkänt och Mycket väl godkänt, inclusive allmänna råd för bedömningsinriktning..." (Skolverket, 2008, p. 21).

All of these documents provided them with a formal framework within which the fourteen texts could be discussed, analyzed and assessed.

The students in this study were in a very different situation. They didn't have access to any explicit criteria whatsoever and as such, their analysis, assessment and grading relied on their accumulated skills and knowledge relating to text analysis and assessment. In this situation they were required to draw on and utilize their own internalized criteria.

I reiterated the points here because despite the poverty of information provided to the students during the study, the data shows that over the span of the fourteen texts, the students employ a common set of assessment criteria in their responses. They utilise all of the assessment criteria ('bedömningsfaktorer') recommended by Skolverket in the teacher's instruction booklet. Further, the number of responses given by the students for each criterion has a strong correlation to those of Skolverket, despite there being areas of deviance (See Diagram 6). The data also shows that while both student and Skolverket responses are weighted in favour of the category Language over Content, students place the most emphasis on Language in their responses.

5.2 Divergence

Both the students and Skolverket have more responses relating to the Language category than the Content category, with the students stressing it most (62% and 55% respectively). In terms of assessment criteria, the major area of divergence in the data between Skolverket and student responses relates to criterion 6 and 11. These criteria fall under the Language category and have to do with understandability and correctness. Proportionately, the number of student responses for these two criteria diverge the most. The criterion relating to Correctness shows the most divergence from Skolverket in terms of percent (10%) and it is also the criterion that has received most responses (29%).

Generally speaking, student responses focus on the weaknesses of texts rather than their strengths. While the student responses are positively focused only 45% of the time, overall Skolverket offers positive responses 70 percent of the time. Further, while Skolverket's responses maintain a constant positive focus for all of the criteria, the focus of the student responses varies. However, despite the negative dominance overall on the student's side, five of the seven criteria have a predominantly positive focus. The negative statistic is primarily generated from one criterion, Correctness, as can be seen in Diagram 9. This fact is reflected in the data which shows student responses relating to the five criteria in the Content category have, like Skolverket, a positive focus.

To sum up, student responses show an increased focus on Language over Content. Within the Language category Understandability and Correctness have comparatively inflated scores. Students are more negative, but that is primarily because of high negative scores for Correctness.

5.3 Grading

Under the column *Exact* in Table 1 is the percentage of students that have the set same grade as Skolverket for the fourteen texts. The other columns show the percentages for the results with the related stress factors included.

(Table 1)

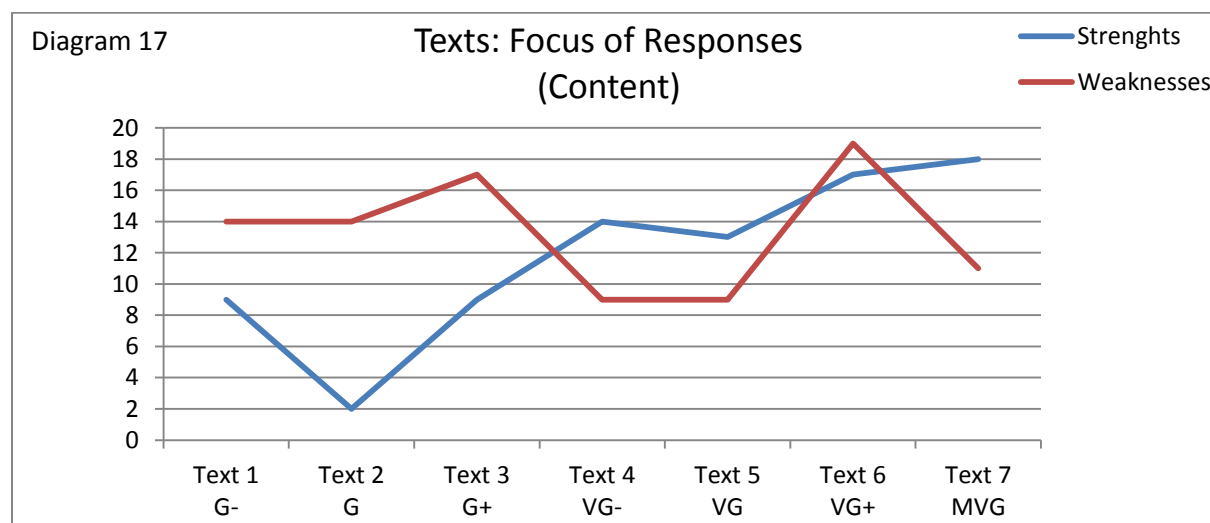
	Stress Factor			
Exact	1/3 of a Grade	2/3 of a Grade	3/3 of a Grade	Whole Grade
24%	54%	73%	88%	60%

As previously documented, the breakdowns for these totals in terms of G, VG and MVG display much higher figures for Godkánt than the total figure and much lower figures for MVG. The relationship is inverse; the higher the grade the lower the accuracy.

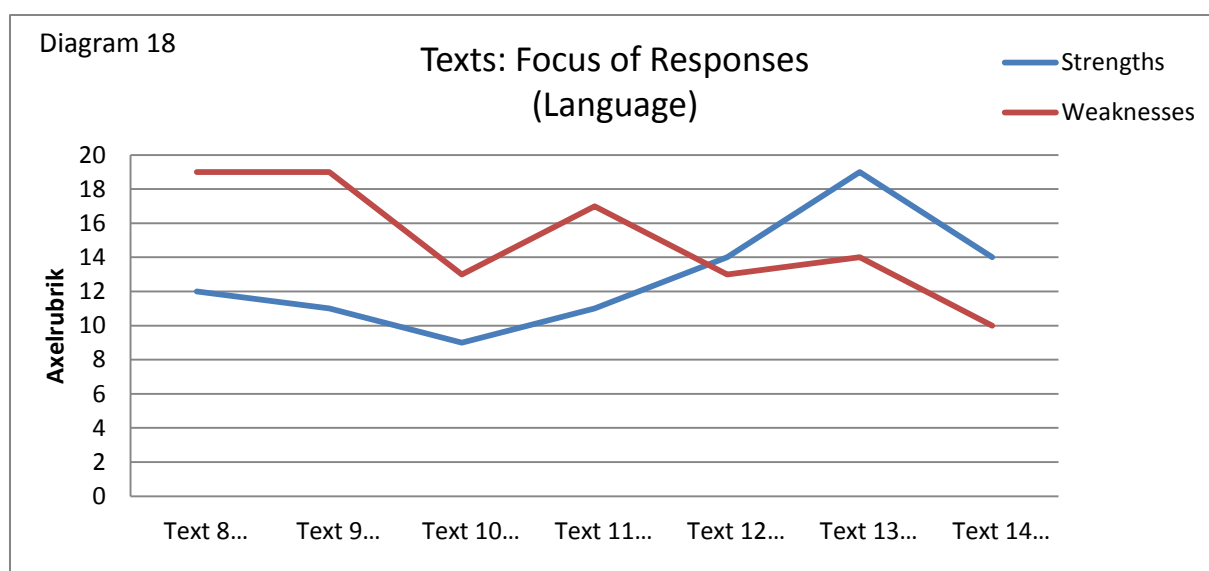
5.4 An Attempt at Synthesis

Is there a link in the data between the student grading and their written responses? As noted in previous graphs, the accuracy of student grading falls sharply from VG-level upwards. As the following graphs indicate, it is precisely at this point that we see a crossover of negative to positive attitudes reflected in students' written responses.

The blue line in Diagram 17 represents the number of positive responses for each text and the red line the negative. As can be seen, aside from one instance relating to text 6, there is a clear transfer at VG—level.



In Diagram 18, the data relating to the Language criteria show a very similar pattern. In this case the crossover point is at VG+ level.



What this seems to indicate is that the students view the more advanced texts more positively. At the VG level students start to offer more positive comments about the texts than negative. The texts Skolverket have graded MVG also receive more positive responses than negative.

The question is, why have the vast majority of the VG and MVG texts been assessed at significantly lower levels by the students than by Skolverket (See Diagram 10 & 11). It would appear that the students are aware that there is a change in quality, but the question is whether can they identify what it is that defines this higher standard.

It would seem appropriate to look for answers in the focus and quality of student responses. Attachment 4 shows the data from the three example texts which Skolverket has assessed to be at MVG level. Consistent with the analysis above, most responses fall under the Language category, in this case 75 percent. This is interesting as this is even marginally higher than the percentage relating for all texts. So it would seem that the quality of the text has little influence on the students' focus on language over content.

Let us now turn to the quality of the responses. Perhaps the best way to get an overview is to look at the text summaries provided for all three texts at MVG level (Table 2). The hash”#” marker signifies Skolverket’s responses. The numbers relate to the different focus groups.

(Table 2)

Text 7	<p>#Helhetsintrycket blir att eleven på ett skickligt sätt uttrycker sig varierat i skrift och kan anpassa framställningen till mottagare och högsta betygssteget.</p> <p>2. En sammanhängande text med få stavfel.</p> <p>3. Uppsatsen var tydligt strukturerad och välskriven. Han lyckades även göra detta ganska tråkiga ämne hyfsat intressant! Ibland kändes inte ord byggnaden och grammatiken helt korrekt.</p> <p>4. Lagom lång och bra text.</p> <p>5. Det är en berättande text med ett hyfsat innehåll dock kunde han försökt att inte använda små ord om och om igen.</p> <p>6. Den var lätt att följa och hade en sammanhängande struktur. Det var ett sämre flyt och hade stundtals lite dålig grammatik.</p>	<p>#Mycket väl godkänt</p> <p>2. VG-</p> <p>3. G+, MVG-</p> <p>4. VG</p> <p>5. VG-, VG, VG, VG</p> <p>6. G+</p>
--------	--	---

Text 13	<p>#Sammanfattningsvis är detta trots allt en text med utmärkt berättarteknik, flyt och anpassning till läsaren.</p> <ol style="list-style-type: none"> 1. Väl uttryckt, men skulle läst igenom texten en gång extra innan inlämning. 2. Många grammatiska fel, kort text men intressant. 3. Bra ordförråd, få stavfel, styckindelning hade underlättat. 4. (Inga kommentarer) 5. Det var en helt okej text, man förstår innehållet men hon kunde ha använt ett mer avancerat ordförråd. 6. Väldigt bra text. 	<p>#Mycket väl godkänt (det nedre skiktet)</p> <ol style="list-style-type: none"> 1. MVG- 2. G+ 3. VG+ 4. VG+ 5. G+, VG- 6. MVG
Text 14	<p>#Over huvud taget är detta en utmärkt text som är synnerligen varierad och skickligt anpassad till såväl syftet som mottagaren.</p> <ol style="list-style-type: none"> 2. Den innehåller vackert språk och ibland ganska avancerat språk. Men en del grammatiska fel och stavfel dessutom fattas det en tydlig inledning och avslutning. Medel bra. 3. Texten var mycket bra överlag förutom några få stavfel och styckindelning som saknades. Moget språk och bra grammatik. 4. Hon ger en bra text med bra flyt och ordförråd. Men gör lite för mycket fel. 5. En text med bra innehåll men med stavfel. 6. Ett bra flyt men man måste läsa nog för att hänga med. 	<p>#Mycket väl godkänt</p> <ol style="list-style-type: none"> 2. VG-, VG 3. MVG- 4. G+, VG, VG 5. VG 6. VG+

I would like to focus on three themes that can be gleaned from the data. The first is that these short summaries of each text vary a great deal in quality. They go from offering very little, “Lagom lång och bra text” to providing a summary of the texts strengths and weaknesses, “Texten var mycket bra överlag förutom några få stavfel och styckindelning som saknades. Moget språk och bra grammatik.” Secondly, I would point out the focus on weaknesses rather than strengths apparent in the comments. As such, these summaries reflect what has been said previously. Thirdly, I would point to the high level of conceptual awareness reflected in many of these comments. Here we see students actively using an advanced meta-language to explain the qualities of texts, with such terms as flyt, styckindelning, och sammanhangande struktur. One group wrote the following: “Uppsatsen var tydligt strukturerad och välskriven. Han lyckades även göra detta ganska tråkiga ämne hyffsat intressant! Ibland kändes inte ord byggnaden och grammatiken helt korrekt.” These comments display the active and effective use of a meta-language for text analysis.

It is difficult to reach a concrete conclusion on the basis of this data alone beyond the fact that the meta-language students employ is varied, often negatively orientated and has a focus on language elements over content.

What is interesting is that the group that I last quoted, in order to illustrate text analysis, set two very different grades for the text in question; G+ and MVG-. Skolverket graded the text as MVG. This group had only two students. The question that arises then is whose thoughts and views have been documented? Clearly they didn’t reach agreement on the grade, as they both chose to record their own. This situation raises a potential shortcoming in this study while at the same time points to its solution.

Could it be the case that the inverse relationship between higher grades and accuracy arises because many students are unable to recognize the qualities of more advanced texts? Vygotsky’s Zone of Proximal Development theory would seem to support to this possibility, as it postulates learning as occurring within set paradigms or zones for individuals

(Verenikina, 2003). Statistically speaking, only around 15 percent of students achieve MVG in English on the National tests (Skolverket, 2007 p. 29). Using this reasoning, only students who already fulfill the goals for MVG can be expected to be able to effectively assess texts at that level.

In closing this analysis section, it may be instructive to recall some points from the literature. While it is important to recognize there is no absolute truth regarding the “correct grade” Skolverket’s assessment is the benchmark for this study. The results in this year’s publication of the study mentioned in Section 2, relating to the auditing of the grading of national tests (Skolinspektionen, 2011) show that 64 percent of marks set by teachers for the Writing section of the Grade 9 English test were the same as those set by Skolinspektionen’s control-testers. Students in my study achieved a very similar match, 60 percent. However, while students generally grade significantly lower than the benchmark, teachers generally grade significantly higher. In summarizing the results relating to teacher assessment, the following conclusion is reached in the report: “Skolinspektionen ifrågasätter, på samma sätt som i förra årets omrättning, huruvida ett av de nationella provens huvudsakliga syften, att stödja en likvärdig bedömning och betygssättning, kan uppnås” (Skolinspektionen, 2011, p. 5).

Skolinspektionen postulates three main reasons for the underlying problem. Firstly, that the test is meant to measure too many things and therefore doesn’t do any of them well. Secondly, that the examples and information provided for teachers to support them in their assessments are too vague, and thirdly, that teachers have trouble being objective with their own students and isolating test performance from other previous performance. All three arguments appear well-founded and in my view, the concept of transparency is of central to all three; not only for teachers, but for students as well. What is needed is a shared system of understanding and a plan which can actively promote it.

6 Conclusions

This study asked a lot of its participants. Without any explicit criteria to work with they were asked to assess and grade texts they had never read before. Nevertheless, the spread of their responses covers the broad range of assessment criteria employed by this study. Furthermore, the number of student responses for each criterion shows a strong correlation with those provided by Skolverket. Put simply, taken over the range of the fourteen texts, the students not only used the same assessment criteria to describe the texts but stressed each criterion to a similar extent in their comments.

A further finding is that student responses have a negative focus. That is to say, they focus more on the weaknesses of texts as opposed to their strengths. Skolverket’s responses have, on the other hand, a very positive focus. However, the negative orientation on the part of the students is largely a result of the fact that the criterion relating to Correctness received more than double the number of responses of any other, and three quarters of student responses for Correctness focus on weaknesses. This means that student responses relating to Correctness (vocabulary, grammar and spelling) generally focus on weaknesses in the texts, and these types of responses heavily influence student replies.

Another finding is that there exists an inverse relationship between students’ accuracy in setting grades and the level of the text they are assessing; the higher the grade the lower the accuracy. Different parameters are examined. The following is an example of the results using one set of parameters. When using a grading system with three levels G, VG and MVG, sixty

percent of all grades set by students correspond to those set by Skolverket. However, if we examine the results in terms of specific grades we find that a full ninety-two percent of all Godkänt grades match Skolverket's, fifty-five percent of Väl godkänt grades and just seventeen percent of Mycket väl godkänt grades.

An analysis of the data shows that below VG-level most of the student responses focus on the weaknesses of texts. At VG-level there is a crossover point after which responses then focus on texts' strengths. This seems to indicate that students recognize the change in quality of texts, however the grading results indicate that accuracy becomes more problematic as the grade-level increases. This raises the question as to whether this crossover is linked to the inverse accuracy of the grading. No conclusions are reached on this point, however a hypothesis is proposed based on reflections arising from this study and grounded in Vygotsky's ZPD theory discussed in Section 2.3.2. It is proposed that the accuracy of assessment and grading may well be compromised if the students doing the assessment are not capable of producing work at a similar or higher level.

In light of the findings a few paths for future research appear fruitful. The first is to examine students' ability to assess texts within their own competency level. An interesting compliment would be to also compare the quality of written comments as well - thereby gaining some idea of conceptual understanding related to the different levels and meta-language usage. Another interesting line of enquiry would be to do a similar study but use formative assessment as a variable. That is, test the assessment practices of students who have actively worked with lots of formative assessment against other groups that haven't. Thirdly, a wider study would also be valuable as it would provide more data for each text and thereby allow analysis of the assessment criteria used in terms of each text, not just for the group of texts as a whole. In conclusion, I would like to add that, in my view, Lgr11 will pose new challenges for both students and teachers in the coming years. Hopefully, a move toward greater transparency will be one of them.

7 References

- Cunningham, G. (1998). *Assessment in the classroom, constructing and interpreting tests*. London and Washington D.C.: The Falmer Press.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. Retrieved 18/9/2011 from <http://www.tandfonline.com.ezproxy.ub.gu.se/doi/abs/10.1080/03075079912331379935>
- Ecclestone, K. (2001) 'I know a 2:a when I see it': Understanding criteria for degree classifications in franchised university programmes, *Journal of Further and Higher Education*, 25:3, 301-313.
- Eisner, E. (1991). Taking a second look: Educational connoisseurship revisited. i M. McLaughlin & D. Phillips (red.), *Evaluation and education at quarter century (pp. 169-187)*. Chicago: The National Society for the Study of Education.
- Frederiksen, J. R., & Collins, A. (1990). A systems approach to educational testing. *CTE Technical Report*, 2, 1-12. Retrieved 20/8/2011 from <http://cct2.edc.org/ccthome/reports/tr2.html>
- Halkier, B. (2010). Focus groups as social enactments: Integrating interaction and content in the analysis of focus group data. *Qualitative Research*, 10: 71-89. Retrieved 5/9/2011 from <http://qrj.sagepub.com/content/10/1/71.short>
- Isemo, A. (2011, 15 november). Nationella prov rättas ofta fel. *Göteborgs posten*, Retrieved 15/11/2011 from <http://www.gp.se/nyheter/goteborg/1.773677-nationella-prov-rattas-ofta-fel>
- Jonsson, A. (2010). The use of transparency in the 'interactive examination' for student teachers. *Assessment in Education: Principles, Policy & Practice*, 17(2), 183-197. Retrieved 19/9/2011 from <http://www.tandfonline.com.ezproxy.ub.gu.se/doi/pdf/10.1080/09695941003694441>
- Kitzinger, J. (1995) Qualitative research: Introducing focus groups. *BMJ*. Retrieved 5/9/2011 from <http://www.bmj.com/content/311/7000/299.full>
- Korp, H. (2003). *Kunskapsbedömning – hur vad och varför*. Stockholm: Myndigheten för skolutveckling.
- Nicol, D., & Macfarlane-Dick, D. (2006) "Formative assessment and self-regulated learning: A model and Sven principles of good feedback practice." *Studies in Higher Education* 31.2, 199-218. Retrieved 27/10/2011 from <http://www.tandfonline.com/doi/pdf/10.1080/03075070600572090>

- O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*, 9(3), 325-335.
Retrieved 18/9/2011 from
<http://www.tandfonline.com.ezproxy.ub.gu.se/doi/pdf/10.1080/1356251042000216642>
- Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, 41(3), 273-290.
Retrieved 12/12/2011 from
<http://www.tandfonline.com.ezproxy.ub.gu.se/doi/pdf/10.1080/14703290410001733294>
- Patel, R. & Davidsson, B. (2003). *Forskningsmetodikens grunder. Att planera, genomföra och rapportera en undersökning* (3:e uppl). Lund: Studentlitteratur.
- Räven, D. and Stephenson, J. (2001) *Competence in the learning society*. New York: Peter Lang.
- Sadler, D. R. (1998) Formative assessment: revisiting the territory, *Assessment in Education*, 5(1), 77-84.
- Shepard, L. (2000). The role of classroom assessment in teaching and learning. *CSE Technincal Report*. Retrieved 25/8/2011 from
http://datause.cse.ucla.edu/DOCS/las_rol_2000.pdf
- Skolinspektionen (2011). Lika eller olika? Omrättning av nationella prov i grundskolan och gymnasieskolan - Redovisning av regeringsuppdrag Dnr. U2009/4877/G. Stockholm: Skolinspektionen.
<http://www.skolinspektionen.se/Documents/omrattning/omratt2011-slutrapport.pdf>
- Skolverket (2000). Grundskola. Retrieved 5/12/2011 from
<http://www3.skolverket.se/ki03/front.aspx?sprak=SV&ar=1112&infotyp=23&skolform=11&id=3870&extraId=2087>
- Skolverket (2005). National assessment and grading in the Swedish school system. Stockholm. Fritzes. Retrieved 5/12/2011 from
<http://www.skolverket.se/publikationer?id=1524>
- Skolverket (2006). Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidshemmet - Lpo 94
Retrieved 2/1/2011 from
<http://www.skolverket.se/publikationer?id=1069>
- Skolverket. (2007). *Provbetyg – Slutbetyg-likvärdig bedömning? sammanfattning av* Stockholm. Fritzes. Retrieved 7/12/2011 from
<http://www.skolverket.se/publikationer?id=1724>
- Skolverket. (2008). English: Grade 9 Bedömning och exempel. Ämnesprov i engelska, äp 9, version 2. Göteborgs universitet: Skolverket.

- Skolverket.(2009). Tolka mål I läroplaner och kursplaner.
Retrieved 2/9/2011 from
http://www.skolverket.se/forskola_och_skola/Grundskoleutbildning/2.716/2.726/2.1526
- Skolverket.(2009a). IUP- processen – Arbetet med den individuella utvecklingsplanen.
Stockholm: Fritzes.
- Skolverket. (2011). Om bedömningsstöd.
Retrieved 2/9/2011 from
http://www.skolverket.se/prov_och_bedomning/ovrigt_bedomningsstod
- Skolverket. (2011a). Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011.
Retrieved 14/12/2011 from
<http://www.skolverket.se/publikationer?id=2575>
- Skolverket. (2011b). Allmänna råd för planering och genomförande av undervisningen
- för grundskolan, grundsärskolan, specialskolan och sameskolan.
Retrieved 12/1/2012 from
<http://www.skolverket.se/publikationer?id=2698>
- Skolverket. (2011c). Kunskapsbedömning i skolan – praxis, begrepp, problem och
möjligheter.
Retrieved 2/1/2012 from
<http://www.skolverket.se/publikationer?id=2660>
- Stake, R. (1999). The goods on American education. I *Phi, Delta, Kappan*, Vol. 80 No. 9 (s. 668-72).
- Stiernstedt, J. (2012, 10 January). ”Sverige ska bli bäst på it”.
Retrieved 10/1/2012 from
<http://www.dn.se/vart-internet/vart-internet---hem/sverige-ska-bli-bast-pa-it>
- Sweden.se. (2009). Education in Sweden.
Retrieved 17/11/2012 from
<http://www.sweden.se/eng/Home/Education/Basic-education/Facts/Education-in-Sweden/>
- Tsagalidis, H. *Yrkeskunandets kinesiska ask*. in; L. Lindström & V. Lindberg (Red.)
Pedagogisk bedömning. Om att dokumentera, bedöma och utveckla kunskap.
Stockholm: HLS Förlag.
- Verenikina, I. (2003). Understanding scaffolding and the zpd in educational research.
Faculty of Education, University of Wollongong, NSW.
Retrieved 20/9/2011 from
<http://www.aare.edu.au/03pap/ver03682.pdf>
- Vetenskapsrådet (2009). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning*. Stockholm: Vetenskapsrådet.
Retrieved 25/8/2011 from

http://www.cm.se/webbshop_vr/pdf/etikreglerhs.pdf

Wiliam and Black (1998). Inside the black box: Raising standards through classroom assessment London: *School of Education*, Kings College.

Retrieved 26/11/2011 from

<http://ww2.fcoe.org/uploads/cgreenlaw/blackbox.pdf>

Wolf, A. (2001) Competence-based assessment. Buckingham: *Open University Press*, 1995.

Retrieved 7/9/2011 from

http://www.heacademy.ac.uk/assets/documents/resources/heca/heca_cl25.pdf

8 Appendix

8.1 Attachment 1

Fokusgrupps arbetsuppgifter

Eleverna sitter i grupper med två till fyra elever. Samtliga grupper får två texter tilldelade och de jobbar med en text åt gången. Alla grupper arbetar med olika texter. Varje grupp ska utse en person som skriver ner gruppens bedömningar.

1. Uppgiften från det gamla nationella provet delas ut till grupperna

Samtidigt som Graeme tydliggör uppgiften och svarar på frågor

2. Text 1 delas ut till respektive grupp

- a) Eleverna läser texten tyst i grupp
- b) Eleverna diskuterar texten i gruppen
- c) Gruppen bedömer texten med hjälp av bifogad mall och skriver ner bedömningen
- d) Gruppen sätter betyg på texten med hjälp av bifogad mall. Graeme samlar in samtliga uppgifter

3. Text 2 delas ut till respektive grupp

Processen är samma som för text 1

Samtliga uppgifter samlas in.

Stort tack för Din medverkan!

Skriftlig bedömning

Exempel text _____

Skriv i punktform med förklaringar

<p>Styrkor med texten (Det som är bra med texten)</p>					<p>Svagheter med texten (Det som behöver utvecklas)</p>				
<p>Hur skulle ni beskriva texten som helhet? (Skriv med meningar)</p>									
<p>Betygsätt texten. (Ringa in det som ni tycker stämmer bäst)</p>									
G-	G	G+	VG-	VG	VG+	MVG-	MVG	MVG+	

Om ni inte kan enas om betyget, skiva ner individuella bedömningar.

Elev 1	Jag ger uppsatsen _____	som betyg _____
Elev 2	Jag ger uppsatsen _____	som betyg _____
Elev 3	Jag ger uppsatsen _____	som betyg _____
Elev 4	Jag ger uppsatsen _____	som betyg _____

Stort tack för Er medverkan!

8.2 Attachment 2

Vänersborg, 2011-10-17

Hej,

Jag heter Graeme Addinsall och jag studerar på Institutionen för pedagogik och didaktik vid Göteborgs universitet. Jag arbetar också som lärare i engelska på Linneaskolan i Ljungskile. Just nu genomför jag ett examensarbete som handlar om bedömning. Syftet med studien är att kartlägga hur eleverna i årskurs 9 bedömer texter i engelska.

Jag planerar att genomföra studien på tre skolor. Elever får i smågrupper läsa två engelska texter (från ett tidigare nationellt prov) och sedan diskutera dem i smågrupper. Detta tar ca 45 min per klass.

Det är självklart frivilligt att delta i undersökningen. Uppgifterna kommer att behandlas enligt gällande etiska principer. Namn på elever eller skola inte kommer att användas. Materialet kommer att behandlas strikt konfidentiellt.

Jag är medveten om att detta tar lite av elevernas dyrbara tid på skolan. Min förhoppning är att övningen blir ett komplement till ordinarie undervisningen då den går ut på att eleverna diskuterar texter från ett tidigare nationellt prov i engelska.

För att elever under 15 år skall få delta i undersökningar krävs det ett samtycke från vårdnadshavare och eleven själv. Genom att underteckna detta dokument ger ni ert samtycke till ovan.

Barnets namn: _____ Underskrift: _____

Målsmans namn: _____ Underskrift: _____

Har ni frågor eller funderingar kring undersökningen är det bara att ringa eller maila mig.

Med vänliga hälsningar,

XXXXXXXXXXXXXX

XXXXXXXXXXXXXX

Tele: XXXXXXXX

Mail: XXXXXXXXXXXXXX

8.3 Attachment 3

Text Summaries		Grade
These are the student responses to the following question on the focus group exercise, "Hur skulle ni beskriva texten som helhet?"		# Skolverket Numbers 1-5 relate to the different focus groups
Text 1	<p>#Trotts stora brister lyckas eleven förmedla en relativ fyllig och levande bild av sitt liv som mopedist. Eleven visa förmåga att berätta och beskriva.</p> <p>2. Bra men störande fel vid grammatik.</p> <p>3. Texten är hyfsat bra och varierande, men pga. alla stavningsfel och grammatiska fel blev det rörigt och svårt att förstå/hänga med.</p> <p>4. (inga kommentarer.)</p> <p>5. Han behöver förbättra mycket. Texten saknar mycket.</p>	#Godkänt (det nedre skiktet) 2. VG- 3. G- 4. G- 5. G-
Text 2	<p>#Trotts de språkliga brister är detta en kommutativt och sammanhängande text som visa viss medvetenhet om mottagaren. Eleven kan förmedla ett innehåll så att budskapet går fram.</p> <p>1. Bra sammanhang, men dåligt uttryck.</p> <p>2. Texten fångade inte upp läsaren.</p> <p>3. Man fatta allt, även om det inte är perfekt grammatik. Uppsatsen har en bra styckindelning, och bra ordförråd. Lite förvirrande stavfel.</p> <p>5. (Inga kommentarer)</p> <p>6 Man förstår vad personen menar och håller sig till ämnet. Men mycket grammatiska fel.</p>	#Godkänt 1. G 2. G- 3. G 5. G, G, G+ 6. G-
Text 3	<p>#Sammanfattningsvis är detta en ge är en engagerad och reflekterande text. Eleven vågar generalisera och problematisera för att förmedla sitt budskap.</p> <p>2. Arbetet var bra, lagom lång, bra ordförråd men några stavfel.</p> <p>3. Texten är väldigt osammanhängande. Den var även svårläst p.g.a. många stavfel.</p> <p>4. Man får en bra inlevelse av texten men det är lite för många stavfel och dåliga formuleringar.</p> <p>5. (inga kommentarer)</p> <p>6. Helt okej men många stavfel.</p>	#Godkänt (det övre skiktet) 2. G+, VG+ 3. G- 4. G 5. G- 6. G+
Text 4	<p>#Sammantaget blir bedömningen att den språkliga säkerheten uppväger bristen på engagemang och anpassning till mottagaren.</p> <p>2. Bra skriven, bra reflektioner och få grammatiska fel.</p> <p>3. Texten var lätt att förstå, den hade en röd tråd och ett bra flyt. Dock hade den några grammatiska- och stavfel.</p> <p>4. Det är en person som skriver om datorer och vad man kan göra på den.</p> <p>5. Bra innehåll, mycket inne rik.</p> <p>6. Personen har utfört uppgiften men behöver utveckla texten.</p>	#Väl godkänt (det nedre siktet) 2. VG+ 3. VG, VG+ 4. G+ 5. G+ 6. G
Text 5	<p>#Framställningen präglas av väl genomfört berättande, [och] är anpassade till mottagaren</p> <p>2. Lite krångligt att förstå, men Annas bra skriven.</p> <p>3. En text med mycket variation och få stavfel. Bra att den som skrev uppsatsen valde ett ämne som den var väldigt passionerad om, men grötig att läsa när det inte var någon styckindelning.</p> <p>4. En helt okej text som behöver slipas.</p> <p>5. Rolig att läsa, Intressant.</p>	#Väl godkänt 2. G+, VG+ 3. VG+, VG- 4. G+, G, G- 5. VG 6. VG

	6. Det var intressant att läsa texten och den hade bra beskrivningar. Dock fanns det några stavfel.	
Text 6	<p>#Sammanfattningsvis är detta en god prestation även om den är något obearbetad både vad gäller innehåll och språk.</p> <ol style="list-style-type: none"> 1. Bra. Fick med mycket. Smått slarvig. 2. Texten var välskriven med tydliga egna åsikter, men med ett flertal felaktiga meningsbyggnader. 3. Texten var väldigt bra skriven och fanns många bra argument. Det var dock väldigt många "I" i texten. 4. (inga kommentarer) 5. Det är en bra text. Texten visar att han/hon kan göra sig förstådd. 6. Texten är försätlig. 	<p>#Väl godkänt (det övre skiktet)</p> <ol style="list-style-type: none"> 1. VG+ 2. VG- 3. VG+, VG- 4. G 5. G 6. G+, VG-, VG-
Text 7	<p>#Helhetsintrycket blir att eleven på ett skickligt sätt uttrycker sig varierat i skrift och kan anpassa framställningen till mottagare och högsta betygssteget.</p> <ol style="list-style-type: none"> 2. En sammanhängande text med få stavfel. 3. Uppsatsen var tydligt strukturerad och välskriven. Han lyckades även göra detta ganska tråkiga ämne hyffsat intressant! Ibland kändes inte ord byggnaden och grammatiken helt korrekt. 4. Lagom lång och bra text. 5. Det är en berättande text med ett hyfsat innehåll dock kunde han försökt att inte använda små ord om och om igen. 6. Den var lätt att följa och hade en sammanhängande struktur. Det var ett sämre flyt och hade stundtals lite dålig grammatik. 	<p>#Mycket väl godkänt</p> <ol style="list-style-type: none"> 2. VG- 3. G+, MVG- 4. VG 5. VG-, VG, VG, VG 6. G+
Text 8	<p>#Helhetsintrycket blir att eleven visar att hon kan ge information och berätta.</p> <ol style="list-style-type: none"> 1. Det skär i vårt akademiska hjärta ... men det är modigt skrivit. 2. I texten finns ingen struktur i texten och stavfelen är tydliga. Personen har dock beskrivit ganska bra om varför den är stolt. 3. Det fanns många stavfel och texten var ostrukturerad. Upprepar samma ord. Det fanns dock mycket info i texten. 4. Tråkig, dålig stavning men håller sig till det den ska skriva om. 5. Texten var dåligt stavat och upplägget kunde varit lite bättre. 6. Han/hon håller sig till uppgiften men gör många fel i stavningen och grammatik. 	<p>#Godkänt (det nedre skiktet)</p> <ol style="list-style-type: none"> 1. G+ 2. G 3. G- 4. G 5. G-, G, G, G 6. G-
Text 9	<p>#Eleven kan ge information på ett begripligt sätt, samt uttrycka och motivera in uppfattning.</p> <ol style="list-style-type: none"> 2. Vi tycker texten har bra ordval med få stavfel, men att texten ofta blir svårförstådd då den innehåller många grammatiska fel. 3. Det verkar som personen har ett litet ordförråd och har svårt att få ihop en riktig bra text. Personen verkar nästan vara rädd för att stava fel. 4. Man förstå den. <p>Den var lätt att läsa, men borde varit längre eftersom eleven fick 80 min att skriva.</p> <ol style="list-style-type: none"> 6. Den var "hackig" eftersom det var typ 60 % punkter. Den var knappt begriplig. 	<p>#Godkänt</p> <ol style="list-style-type: none"> 2. G 3. G 4. G-, G, G 5. G 6. G-
Text 10	<p>#Sammantaget blir intrycket att detta är en strukturerad och engagerande text med gott språkligt flyt.</p> <ol style="list-style-type: none"> 2. En mycket bra genomtänkt text, men längre borde den varit. 3. Kort välskriven text med bra grammatik och nästan inga stavfel. Kan jobba på att utveckla texten mer för att få den både längre och mer intressant. 4. Det var ett barn som var stolt över sin pappa. 5. Den var förståeligt och kortfattat, och ett okej innehåll. 	<p>#Godkänt (det övre skiktet)</p> <ol style="list-style-type: none"> 2. VG 3. G+ 4. G 5. G

Text 11	<p>#Helhetsintrycket blir att detta är en läsvärd och reflekterande text som är språkligt något ojämn men mycket kommunikativ.</p> <p>2. Berätta mycket och håller sig till ämnet. Jättebra text.</p> <p>3. Hade bra koll på skrivreglerna. Det var lagom lång text.</p> <p>4. Bra men stakig.</p> <p>5. Bra ämne med bra innehåll och inga stavfel, men kunde varit bättre upplägg.</p> <p>6. Ingen rolig text.</p>	<p>#Väl godkänt (det nedre skiktet)</p> <p>2. MVG-</p> <p>3. VG-</p> <p>4. G+</p> <p>5. VG, VG+, VG+</p> <p>6. G, G, G+</p>
Text 12	<p>#Helhetsintrycket blir att detta är en mycket personlig och färgstark text med gott flyt och varierat språk.</p> <p>2. Hon framhäver att hon kan stå för sina åsikter. Hon vet vad som är rätt och fel och faller inte för grupptrick.</p> <p>3. Texten var lätt att läsa och hyfsat välskriven. Det innehåll några stavfel som inte påverkade texten.</p> <p>5. Texten är uppbyggande. Hon skriver vad hon kan. Stadig och bra text.</p> <p>6. Det är en mycket bra text och trots stavfel förstår man allt ändå.</p>	<p>#Väl godkänt (det övre skiktet)</p> <p>2. G+</p> <p>3. VG</p> <p>5. G+, VG-</p> <p>6. VG</p>
Text 13	<p>#Sammanfattningsvis är detta trots allt ett Text på en text med utmärkt berättarteknik, flyt och anpassning till läsaren.</p> <p>1. Väl uttryckt, men skulle läst igenom texten en gång extra innan inlämning.</p> <p>2. Många grammatiska fel, kort text men intressant.</p> <p>3. Bra ordförråd, få stavfel, styckindelning hade underlättat.</p> <p>4. (Inga kommentarer)</p> <p>5. Det var en helt okej text, man förstår innehållet men hon kunde ha använt ett mer avancerat ordförråd.</p> <p>6. Väldigt bra text.</p>	<p>#Mycket väl godkänt (det nedre skiktet)</p> <p>1. MVG-</p> <p>2. G+</p> <p>3. VG+</p> <p>4. VG+</p> <p>5. G+, VG-</p> <p>6. MVG</p>
Text 14	<p>#Over huvud taget är detta en utmärkt text som är synnerligen varierad och skickligt anpassad till såväl syftet som mottagaren.</p> <p>2. Den innehåller vackert språk och ibland ganska avancerat språk. Men en del grammatiska fel och stavfel dessutom fattas det en tydlig inledning och avslutning. Medel bra.</p> <p>3. Texten var mycket bra överlag förutom några få stavfel och styckindelning som saknades. Moget språk och bra grammatik.</p> <p>4. Hon ger en bra text med bra flyt och ordförråd. Men gör lite för mycket fel.</p> <p>5. En text med bra innehåll men med stavfel.</p> <p>6. Ett bra flyt men man måste läsa noga för att hänga med.</p>	<p>#Mycket väl godkänt</p> <p>2. VG-, VG</p> <p>3. MVG-</p> <p>4. G+, VG, VG</p> <p>5. VG</p> <p>6. VG+</p>

8.4 Attachment 4

This attachment contains the data for all three texts Skolverket has assessed as at MVG level. The hash symbol “#” signifies Skolverket’s comments. The numbers relate to different focus groups.

Example 7: In a World Full of Things	Focus on Strengths	Focus on Weaknesses
Innehåll		
Om texten ger uttryck för en vilja att använda språket för att förmedla ett innehåll	#Det här är en mycket innehållsrik och väldisponerad text. #Eleven beskriver, förklarar, diskuterar och jämför. 5. Det är en bra berättande text.	3. Ämnet intresserade oss inte jätte mycket.
Om texten är sammanhängande och strukturerad	3. Tog upp både för- och nackdelar. 3. Bra indelad i stycken. 3. Lagom lång 3. Ganska tydlig röd tråd. 4. Höll sig till ämnet.	#Möjligen skulle texten kunnat få en tydligare avrundning 2. Dålig styckindelning. 2. konstig struktur av text. 3. Ordföljden kändes lite konstig ibland. 4. Styckindelningar.
Om innehållet är fylligt eller magert/torftigt	#Känslouttryck och förstärkningsord gör innehållet engagerande och personligt 4. Väl förklarat.	
Om ämnet är utförligt behandlat - om eleven fokuserar eller bara ytligt behandlar ämnet/de olika punkterna		
Om texten är anpassad till mottagaren/syftet	#Tonen är eftertänksamt resonerande #Den känns äkta och väl anpassad till både mottagaren och syftet. 6. Håller sig till ämnet.	
Språk		
Begriplighet – förmåga att uttrycka ett budskap klart och tydligt	3. Bra skriven. Tog allt från början. 3. Man förstå allt, varför han hade fått sin moped och varför han gillade den så mycket bl. a.	

	3. Han verkar vara en bra förare och ta det lugnt i trafiken. Han förstå även farorna med en moped.	
Ledighet, variation och säkerhet – flyt	#Texten har genomgående ett utmärkt flyt i präglas av korrekthet.	6. Dåligt flyt.
Strategier att ta sig runt språkliga problem		
Vokabulär och idiomatik (omfång, variation)	#Eleven behärskar ganska avancerade grammatiska strukturer och verbfraser #Ordförrådet är stort 2. Bra ordförråd. 6. Bra ordförråd.	#Enstaka "svengelska" misstag förekommer 5. För lätta ord. 5. Samma ord om och om igen.
Meningsbyggnad – förmåga att binda samman satser och meningar	#Men den är annars klart strukturerad med väl genomförd styckesindelning.	
Korrekthet (vokabulär, idiomatik, grammatik och stavning)	#språket är ledigt, idiomatiskt och ger ett säkert intryck. 2. Få stavfel. 3. Få stavfel 4. Inga grammatiska fel	2. Många grammatiska fel 4. Skrev "motor" istället för "engine". 6. Rätt dålig grammatik ibland.
Helheten	#Helhetsintrycket blir att eleven på ett skickligt sätt uttrycker sig varierat i skrift och kan anpassa framställningen till mottagare och högsta betygssteget. 2. En sammanhängande text med få stavfel. 3. Uppsatsen var tydligt strukturerad och välskriven. Han lyckades även göra detta ganska tråkiga ämne hyffsat intressant! Ibland kändes inte ord byggnaden och grammatiken helt korrekt. 4. Lagom lång och bra text. 5. Det är en berättande text med ett hyfsat innehåll dock kunde han försökt att inte använda små ord om och om igen. 6. Den var lätt att följa och hade en sammanhängande struktur. Det var ett sämre flyt och hade stundtals lite dålig grammatik.	
Bedömningen:	#Mycket väl godkänt 2. VG- 3. G+, MVG- 4. VG 5. VG-, VG, VG, VG 6. G+	

Example 13: Proud Of...	Focus on Strengths	Focus on Weaknesses
Innehåll		
Om texten ger uttryck för en vilja att använda språket för att förmedla ett innehåll		
Om texten är sammanhängande och strukturerad	#kronologiskt uppbyggd och väl inramad IV inledning och avslutning 2. Styckindelning 3. Bra början.	1. Rörig text 2. Ingen inledning 3. Ingen styckindelning 3. Jobba på slutet
Om innehållet är fylligt eller magert/torftigt		3. Jämför mer. 5. Texten borde ha varit längre för ett högre betyg.
Om ämnet är utförligt behandlat - om eleven fokuserar eller bara ytligt behandlar ämnet/de olika punkterna		
Om texten är anpassad till mottagaren/syftet	#Läsarens förväntan byggs upp genom flera exempel på utmaningar och känslouttryck. Ett par målande bilder, blir också effektfulla. 6. Allt annat är utmärkt.	
Språk		
Begriplighet – förmåga att uttrycka ett budskap klart och tydligt	4. Texten är lätt att förstå. 5. Man förstår vad det handlar om.	
Ledighet, variation och säkerhet – flyt	#Här får läsaren ta del av en flödande berättelse 1. Bra beskrivningen 2. Bra berättelse. 3. Bra berättat 3. Förklara bra 3. Beskriva bra. 4. Bra beskrivningar. 5. Okej beskrivningar.	
Strategier att ta sig runt språkliga problem		

Vokabulär och idiomatik (omfång, variation)	<p>#ordförrådet räcker väl för att behandla det valda ämnesområdet</p> <p>#Eleven uttrycker sig relativt idiomatiskt</p> <p>2. Ganska bra ordförråd</p> <p>4. Bra och varierande ordförråd.</p>	<p>#[Ordförrådet] är inte så avancerat.</p> <p>2. Uppprepning av competition.</p> <p>2. Glömt ord.</p> <p>5. Mer avancerat ordförråd.</p>
Meningsbyggnad – förmåga att binda samman satser och meningar	<p>#Meningarna är mycket smidigt sammanbundna</p> <p>1. Meningsbyggnaden</p>	
Korrekthet (vokabulär, idiomatik, grammatik och stavning)	<p>#Verbbehandlingen är god.</p> <p>1. Stavningen</p> <p>3. få stavfel</p> <p>5. Bra grammatik</p> <p>6. Inga stavfel.</p>	<p>#del förväxlingar när det gäller stavning men också grammatik förekommer.</p> <p>#[Ordförrådet] influeras ibland av modersmålet</p> <p>1. Lite stavfel</p> <p>1. Svengelska</p> <p>2. Många grammatiska fel</p> <p>2. Saknar kommatecken och punkter.</p> <p>3. Några grammatiska fel</p>
Helheten	<p>#Sammanfattningsvis är detta trots allt ett exempel på en text med utmärkt berättarteknik, flyt och anpassning till läsaren.</p> <p>1. Väl uttryckt, men skulle läst igenom texten en gång extra innan inlämning.</p> <p>2. Många grammatiska fel, kort text men intressant.</p> <p>3. Bra ordförråd, få stavfel, styckindelning hade underlättat.</p> <p>4. (Inga kommentarer)</p> <p>5. Det var en helt okej text, man förstår innehållet men hon kunde ha använt ett mer avancerat ordförråd.</p> <p>6. Väldigt bra text.</p>	
Bedömningen:	<p>#Mycket väl godkänt (det nedre skiktet)</p> <p>1. MVG-</p> <p>2. G+</p> <p>3. VG+</p> <p>4. VG+</p> <p>5. G+, VG-</p> <p>6. MVG</p>	

Example 14: Proud Of...	Focus on Strengths	Focus on Weaknesses
Innehåll		
Om texten ger uttryck för en vilja att använda språket för att förmedla ett innehåll	#Argumentationen fördjupas eftersom eleven genomgående diskuterar såväl omständigheter som konsekvenser.	
Om texten är sammanhängande och strukturerad	#Detta är en mycket väl sammanhållen text med tydlig tankestruktur, logiska övergångar och avslutningsvis en fin återkoppling till inledningen. 4. Det är en bra text.	2. Inte särskilt bra struktur utan någon tydlig inledning eller tydligt avslut. 3. Ingen styckindelning 6. Kunde varit mer sammanhängande. Det hoppade.
Om innehållet är fylligt eller magert/torftigt	#Innehållet blir engagerande, fylligt och informativt.	
Om ämnet är utförligt behandlat - om eleven fokuserar eller bara ytligt behandlar ämnet/de olika punkterna	#Ämnet behandlas mycket utförligt.	
Om texten är anpassad till mottagaren/syftet	#Eleven skriver om flera personer och lyckas hålla fokus på vars och ens beundransvärda framåtanda.	
Språk		
Begriplighet – förmåga att uttrycka ett budskap klart och tydligt	#Språkligt sett ger texten ett mycket säkert och varierat intryck. 2. Det är fint skrivit och texten innehåller ganska avancerat språk. 3. Moget språk 5. Förstod innehållet. Bra innehåll.	
Ledighet, variation och säkerhet – flyt	3. Lagom lång 4. Bra flyt. 6. Bra flyt 6. Varierande	
Strategier att ta sig runt språkliga problem	#Eleven tänjer sitt språk, använder strategier och vågar ta risker.	#Detta leder också till att han ibland gör misstag, men de kan inte sägas störa det goda helhetsintrycket nämnvärt.

Vokabulär och idiomatik (omfång, variation)	#Eleven behärskar en rik vokabulär och avancerad syntax Flera av uttrycken skapar detaljrika och färgstarka bilder hos läsaren 3. Bra ordförråd 3. Bra grammatik Bra stavning 5. Svåra ord 6. Många olika ord.	
Meningsbyggnad – förmåga att binda samman satser och meningar	#meningsbyggnaden varierad med t.ex. retoriska frågor och korta effektfulla påståenden 2. Bra meningsbyggnad	5. Fel ordning
Korrekthet (vokabulär, idiomatik, grammatik och stavning)	#verbbehandlingen är genomgående god 2. Bra användning av kommatecken oftast.	2. En hel del grammatiska fel och små stavfel. 4. Några små stavfel. 5. Stavfel 5. Grammatik fel 6. Något litet grammatisk fel 6. Några få stavfel.
Helheten	#Over huvud taget är detta en utmärkt text som är synnerligen varierad och skickligt anpassad till såväl syftet som mottagaren. 2. Den innehåller vackert språk och ibland ganska avancerat språk. Men en del grammatiska fel och stavfel dessutom fattas det en tydlig inledning och avslutning. Medel bra. 3. Texten var mycket bra överlag förutom några få stavfel och styckindelning som saknades. Moget språk och bra grammatik. 4. Hon ger en bra text med bra flyt och ordförråd. Men gör lite för mycket fel. 5. En text med bra innehåll men med stavfel. 6. Ett bra flyt men man måste läsa noga för att hänga med.	
Bedömningen:	#Mycket väl godkänt 2. VG-, VG 3. MVG- 4. G+, VG, VG 5. VG 6. VG+	